university of groningen

University Medical Center Groningen

University of Groningen

Disability in patients with chronic neck pain

Jorritsma, Wim

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Jorritsma, W. (2013). *Disability in patients with chronic neck pain: the clinimetric properties of instruments*. [Thesis fully internal (DIV), University of Groningen]. [S.n.].

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 05-05-2023

Disability in patients with chronic neck pain

The clinimetric properties of instruments

Wim Jorritsma

Disability in patients with chronic neck pain

The clinimetric properties of instruments

Wim Jorritsma

Stellingen

behorende bij het proefschrift

Disability in patients with chronic neck pain

The clinimetric properties of instruments

1. De Nederlandse versies van de Neck Pain and Disability Scale (NPAD) en de Neck Disability Index (NDI) zijn valide meetinstrumenten van zelf-gerapporteerde nekpijn gerelateerde 'disability' bij patiënten met aspecifieke chronische nekpijn. (*dit proefschrift*)
2. De responsiviteit van de NPAD en NDI bij patiënten met aspecifieke chronische nekpijn na een multidisciplinair pijnrevalidatieprogramma is gelijk. (*dit proefschrift*)
3. De Physical Dysfunction Severity (PDS) en de modified cervical Non Organic Signs (mcNOS) kunnen relevante informatie geven in de biopsychosociale diagnostiek van patiënten met aspecifieke chronische nekpijn. (*dit proefschrift*)
4. De Short Form-36 Health Survey (SF-36) heeft een goede constructvaliditeit en kan worden gebruikt om zelf gerapporteerde algehele gezondheid te meten bij patiënten met aspecifieke chronische nekpijn in een universitaire revalidatiepolikliniek. (*dit proefschrift*)
5. De NPAD, NDI, PDS, mcNOS, SF-36 en de functionele capaciteitsevaluatie (FCE) meten elk andere dimensies van het multidimensionele construct 'disability' bij patiënten met aspecifieke chronische nekpijn. (*dit proefschrift*)
6. 'De' patiënt met nekpijn bestaat niet.
7. Segmentale neurofysiologische reacties op een functiestoornis van het bewegingsapparaat zijn vaak een onbegrepen fenomeen voor patiënten en medici.
8. Wanneer basisartsen in hun opleiding meer kennis zouden opdoen over de functionele morfologie en de effecten van training en immobilisatie, zou dat de diagnostiek en behandeling van klachten van het bewegingsapparaat verbeteren.
9. Het door clinici toeschrijven van pijnklachten van patiënten aan 'scheefstanden', 'verkalking', 'ontkalking' en 'slijtage' draagt bij aan catastroferende cognities van de patiënt.
10. Om osteoporose te voorkomen kan men beter met de koe uit wandelen gaan dan haar melk drinken. (*vrij naar Walter Willett 2011*)
11. Aanstaande regeringsleiders kunnen het vertrouwen van de kiezers in hun capaciteiten vergroten door de snelheid waarmee ze traplopen op te voeren.
12. Learning by teaching is a joy forever.
13. La cucina Italiana is het beste exportprodukt van Italië.
14. Het feit dat hoger opgeleiden hun vaatdoekjes langer gebruiken voordat ze gewassen worden en een hogere levensverwachting hebben, vraagt nader onderzoek.

Wim Jorritsma

Groningen, 19 juni 2013

Printing of this thesis was generously supported by:
- Stichting Manuele Geneeskunde, Eindhoven, The Netherlands
- Ontwikkelcentrum Pijnrevalidatie, Centrum voor Revalidatie UMC Groningen, The Netherlands
- Stichting Beatrixoord Noord Nederland, Groningen, The Netherlands

Disability in patients with chronic neck pain
The clinimetric properties of instruments

Jorritsma, W

Dissertation University of Groningen, the Netherlands – With ref. – With summary in Dutch

Cover	Rutger Hiemstra, www.rutgerhiemstra.com
Layout	Renate Siebes, Proefschrift.nu
Printed by	Ridderprint, Ridderkerk
ISBN	978-90-367-6239-7

© W. Jorritsma, Paterswolde, The Netherlands 2013
All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronical or mechanical, including photocopy, recording or any information storage or retrieval system, without the prior written permission of the copyright owner.

RIJKSUNIVERSITEIT GRONINGEN

Disability in patients with chronic neck pain

The clinimetric properties of instruments

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
woensdag 19 juni 2013
om 12.45 uur

door

Wim Jorritsma

geboren op 10 december 1947
te Lochem

Promotores	Prof. dr. J.H.B. Geertzen Prof. dr. M.F. Reneman Prof. dr. P.U. Dijkstra
Beoordelingscommissie	Prof. dr. H.J. Arendzen Prof. dr. S.K. Bulstra Prof. dr. B.W. Koes

Ter nagedachtenis aan mijn vader

Voor Lideke
Annelies, Hans en Reinder

CONTENTS

Chapter 1

General introduction

1. NECK PAIN

1.1. Epidemiology

Neck pain is one of the most common musculoskeletal disorders [35, 78] with an estimated point prevalence in the general population between 10% and 22% [10, 13, 26, 78]. The 1-year prevalence of neck pain is estimated between 32% and 36% of adults in the general population [13, 26, 77, 78]. The 1-year prevalence of chronic (>3 months duration) neck pain (CNP) in the general population is between 18% to 21% [13, 78]. About two-thirds of the population will experience neck pain at some point in their life [7]. Prevalence rises with age and is the highest in the age group between 40–59 years and thereafter decreases slightly [7, 12]. Women are affected almost twice as much as men [7, 12, 78].

1.2. Pathology

Neck pain may arise from different structures in the neck such as muscles, ligaments, intervertebral discs, zygapophyseal joints, dura and neural tissue [9]. There are a large number of specific causes of neck pain such as radiculopathy, systemic rheumatic disorders, fractures, infections, (spinal) tumors, inflammation, thyroid disorders, oesophageal obstruction or reflux disease and congenital disorders [9, 35]. In the large majority of patients suffering from neck pain however, no specific underlying pathology can be established and the neck pain is labeled as 'non-specific' or 'mechanical' because it is produced or aggravated by neck movements or sustained neck postures [9, 35]. In this group of patients radiologic degenerative changes are usually not found or occur in the same frequency as in persons without neck complaints [9].

1.3. Risk factors

Neck pain is assumed to be of multi-factorial origin. A number of individual, physical and psychosocial risk factors are associated with the onset of neck pain.

Individual risk factors are female gender, age, smoking, exposure to environmental tobacco and stressful life events [19]. An important physical risk factor is prolonged static loading of muscles, zygapophyseal joints, ligaments and disci in a slumped (sitting) posture with too few breaks, which may lead to "tension neck syndrome" [94]. Constrained head and arm posture, repetitive arm and neck movements may increase physical loading of different tissues. For work-related physical risk factors there is some evidence for a positive association between neck flexion posture, arm force, arm posture, duration of sitting, twisting or bending the trunk, hand-arm vibration and workplace design and the onset of neck pain [3].

Psychosocial factors may cause stress which may increase biomechanical strain due to higher tension of the neck/shoulder muscles and also lead to physiological neural and hormonal reactions.

Poor psychological health (mental distress, any mental disorder, depression, depressed mood and anxiety) and coping styles (avoidance behavior and social support seeking) are risk factors for the onset of neck pain, but factors related to compensation, laws or society are not reported as risk factors in the general population [19, 42]. For work-related psychosocial risk factors there is some evidence for a positive association between high quantitative job demands (e.g. working under time pressure or working with deadlines), low social (co-worker) support, low job control, high and low skill discretion and low job satisfaction and the onset of neck pain [4, 18].

1.4. Pain, disability and natural course

Although most people with neck pain generally report only mild pain and mild disability due to neck pain [34], 5% to 10% report severe disability [24, 28]. Older individuals report higher scores of pain intensity and more neck pain related disability compared to younger individuals [34].

The natural course of neck pain is characterized by exacerbations and remissions [26]. Between 20% and 50% of individuals with neck pain in the general population and general practice will develop CNP [13, 26, 105]. Only about a third of the individuals in the general population and in general practice with prevalent neck pain at baseline will have complete resolution of it within one year [12, 19, 105]. Of those who experience neck pain at some point, between 50% and 85% will report neck pain again, 1 to 5 years later [39].

1.5. Prognostic factors

Prognostic factors can potentially predict the clinical course (recovery or failure to recover) after onset of non-specific neck pain. Prognostic factors for a poor outcome in the general population are lower social class, poor self-perceived general health, older age (>45 years), catastrophising, anxiety, initial pain intensity, low treatment expectations, initial disability severity [26, 41]. Most prognostic factors which predict poor outcome in persons with neck pain are the same as risk factors for neck pain [19].

Prognostic factors for a poor outcome in the Dutch primary care setting are high severity of physical neck dysfunction during physical examination, higher age, traumatic cause of neck pain, severity of neck pain, concomitant low back pain and headache, previous episodes of neck pain, fear avoidance behavior, severity of baseline disability [45, 81, 89]. A score chart based on a prediction model was developed in a primary care setting to estimate the probability of persistent complaints at 6 months in non-specific neck pain [89].

1.6. Who seeks care?

In the general population between 15% and 40% of the individuals with neck pain will consult their general practitioner annually [2, 25, 78]. Various factors such as neck pain severity, duration,

socio-economic status and presence of comorbidity are predictors for those who seek care [25]. Patients with neck pain due to a car collision and patients who attributed neck pain to an occupational factor were more likely to seek help than patients with neck pain that occurred in other settings [25]. In the Dutch general practice the reported incidence density and consultation rate for neck pain were much higher for patients with public health insurance than for privately insured patients, which is probably related to the socio-economic status [12]. More strenuous working conditions and/or poorer health may contribute to an increased consultation rate for the group with a public health insurance [12]. Chronic pain patients who consulted a tertiary pain centre differed with regard to important psychological characteristics from patients with chronic pain lasting >1 year and who had not consulted a physician because of the pain during the last year [85]. The health care seekers had higher levels of distorted cognition (such as catastrophising and negative self-efficacy), more pain related distress (such as interference and impediment due to pain, mood affected by pain), lower general activity levels and lower levels of internal locus of control [85].

1.7. Sick leave and total costs

Although neck pain is not life-threatening, it may affect daily-life activities and participation in work. In the Dutch general population the 1-year prevalence of sick leave due to pain of neck, shoulder or higher back was 28%; of which 8% was less than 1 week, 8% was between 1 and 4 weeks, 6% was more than 4 weeks and for 6% it was not applicable [78]. In the Dutch general practice between 28% to 35% of the patients with acute neck pain was on sick leave due to neck pain [104, 105]. Almost half of these patients returned to work within 7 days. After 1 year self-reported periods of sick leave were up to 1 week for 37%, between 1 week and 1 month for 22%, between 1 and 3 months for 20% and above 3 months for 21% of patients [105]. In the aforementioned study 8% of the patients were still on sick leave at the 1-year follow up.

Sick leave has a large impact on the total costs of neck pain in the Netherlands. Approximately 1% of total health care expenditure is utilized for the treatment of neck pain [10]. Direct medical costs were estimated to be 23% of the total costs and the paramedical care accounted for the largest portion (84%) [56]. The majority (77%) of total costs were estimated to be indirect non-medical costs due to sick leave (27%) and disability payment (50%) [56].

2. CLINICAL OBSERVATION

2.1. Clinical history

As for any pain problem a comprehensive history of neck pain can be obtained by interviewing patients about main symptoms, duration of symptoms, mode, circumstances and time of onset, site and radiation of pain, quality of pain, if symptoms are constant (every second of the day)

or intermittent, if specific activities precipitate, aggravate or diminish the symptoms, associated symptoms and if there are prior episodes of neck pain. Neck pain may be accompanied by a subjective feeling of stiffness, limited range of motion of the neck, numbness, tingling or weakness in the upper limbs, dizziness, dysphagia, dyspneu or triggering of migraine [8].

Referred pain according to dermatomal and myotomal patterns from the neck to the head, between the shoulders, to the dorsal scapular region, to the arm or to the anterior chestwall is very common [9, 35]. For pain originating from the spinal muscles, cervical joints and discs maps of cervical referred pain distribution have been made, based on observations in volunteers [9, 22]. The referred pain distribution is quite similar for each structure because of segmental innervation. Pain from the upper cervical structures is referred towards the head and from the lower cervical structures to the upper limb girdle [9, 22]. Additionally, the neck is a site for referred pain from cardiac, lung, gastric and diaphragmatic diseases [35]. Generally referred pain is perceived deeply and aching or pressure-like in quality while radicular pain is sharp, shooting or lancinating and travels along the arm in a narrow band [9]. In patients with mechanical neck pain and patients with cervical radiculopathy, components of the clinical history (main symptoms, mode of onset of the symptoms, if symptoms are constant or intermittent, if specific activities aggravate or diminish the symptoms and if there are prior episodes of neck pain) have an interobserver reliability ranging from Kappa (K) = 0.67 to K = 0.90 [21, 108].

In patients with CNP concomitant symptoms were reported such as frustration and fatigue (50%), depressive mood, anger and dizziness (35%), concentration problems (30%), memory loss and low back pain (25%) and nausea (20%) [32, 44]. Disability affecting activities of daily life such as interferences with sleep, driving a car, lifting, working overhead, looking overhead, housekeeping, gardening may be present [100, 111]. Participation in work for wages, sports and socializing may be restricted [100, 111].

It is important to identify "red flags" (symptoms for potentially serious specific conditions) in patients with neck pain (Table 1.1) [8, 72, 74, 87].

Besides identifying red flags it is of interest to identify 'yellow', 'blue', 'black' and 'orange flags' [53, 72, 92, 106]. Previously, yellow flags were psychosocial risk factors for prolonged disability and failure to return to work such as attitudes and beliefs, experienced emotions and mood, diagnostic and treatment issues, work problems, compensation issues and litigation [53] (Table 1.2).

In recent years the 'yellow flags' have rather more been reserved for more overtly psychological risk factors and workplace risk factors are divided into 2 categories: 'blue flags' and 'black flags' [64]. Blue flags are the individual worker's perceptions about the relationship between work and health [62, 64, 73]. They are conceptualized as workers' perceptions of a stressful, unsupportive, unfulfilling or highly demanding work environment (Table 1.3).

Table 1.1 Red flags[a] in clinical history of neck pain

History	Symptoms
• acute neck pain with (violent) trauma • neck pain without clear etiology and history of cancer • neck pain with inflammatory disorders (ankylosing spondylitis, rheumatoid arthritis) suspected • history of severe osteoporosis or systemic steroid use • drug abuse, human immune deficiency, virus infection	• neck pain with unexplained weight loss, no appetite, distaste of meat and/or alcohol • neck pain with parasthesia in arms and legs, clumsy hands and gait disturbance • neck pain with sharp, shooting or lancinating pain and pins and needles in the arm with certain movements of the neck • neck pain with fever, night sweats • neck pain with arthritis symptoms and signs in peripheral joints elsewhere • drop attacks especially when moving the neck • neck pain in children (<18 year) with considerable pain • intense pain on minimal motion of the neck • constant progressive pain at night

[a] Red flags are symptoms and/or signs for potentially serious specific conditions.
History and symptoms compiled and slightly modified from Nachemson and Vingård, 2000; Binder, 2007; Rubinstein and Van Tulder, 2008; Nordin et al., 2009.

Table 1.2 Yellow flags[a] in clinical history of neck pain

Cognitions about neck pain and treatment	***Coping style with neck pain***
• neck pain is harmful or disabling • catastrophising (thinking the worst) • misinterpreting bodily symptoms • belief that neck pain is uncontrollable • experience of catastrophising diagnoses or explanations for neck pain • expectation that passive treatments rather than active participation will help • belief that all pain must have ceased before attempting to return to normal activities or work	• avoiding neck rotation or extension movements or activities due to misplaced anticipation of pain (fear avoidance behavior) • reduced activity levels • progressive substitution of life style away from productive activity • use of extended rest (disproportional downtime) • excessive reliance on use of aids or appliances
Experienced emotions/mood	***Social issues***
• tendency to low mood • fear of increased pain with activity or work • more irritable than usual • heightened awareness of body sensations • feeling under stress and unable to maintain sense of control	• withdrawal from social interaction • family/relation problems • work-related problems • history of extended time off work due to neck pain or other pain problem (e.g. more than 12 weeks) • history of previous neck pain with a previous claim and time off work • low educational background, low socioeconomic status • financial problems

[a] Yellow flags are psychosocial risk factors for prolonged disability and failure to return to work.
Adapted from Kendall et al., 1997 (Guide to assessing psychosocial yellow flags in acute low back pain risk factors for long time disability and work loss).

Black flags are observable workplace characteristics and nature of the work (specific for the company or industry), as well the insurance and compensation system under which workplace injuries are managed [64]. Orange flags are clearly abnormal psychological or psychiatric factors or disorders (e.g. posttraumatic stress disorder, major depression), suggestive of diagnosable psychopathology [64].

2.2. Physical examination

The basic elements of the physical examination of the musculoskeletal system include inspection, active and passive movements, palpation of tender points and a neuromuscular examination to evaluate tendon reflexes, muscle strength, distribution of any sensory complaint and neurological provocative tests of the neck [67, 72, 87]. Clinicians trained in manual skills such as manual therapists, physicians in musculoskeletal medicine and chiropractors might also examine inter-segmental motion and apply segmental pain provocation tests [21, 80]. Clinicians should be alert for signs with a high probability of being associated with specific neck disorders by making use of red flags in physical examination (Table 1.4) [8, 72, 74, 87].

Table 1.3 Blue flags[a] in clinical history of neck pain

- work history, including patterns of frequent job changes
- lack of job satisfaction (unfulfilling work environment)
- high quantitative job demands (e.g. working under time pressure or working with deadlines)
- belief that work is harmful; that it will do damage or will be dangerous (fear of re-injury)
- low expectations of returning to work
- belief that workplace supervisor and workmates are unsupportive
- lack of vocational direction

[a] Blue flags are the individual worker's perceptions about the relationship between work and health.
Adapted from Main and Burton 2000 and Shaw et al., 2009.

Table 1.4 Red flags[a] in physical examination of the neck

- any region of swelling or neck lumps (supraclavicular, thyroid, lymphadenopathie)
- exquisite tenderness over vertebral spinous processus
- severe restriction of cervical flexion
- sharp shooting pains in shoulders, arms or upper back with certain movements of the neck
- diminished or absent reflex
- objective muscle wasting or weakness
- sensory loss of a clear dermatomic distribution
- hyperreflexia with positive Babinski
- clumsy hands and/or gait disturbance
- numbness or tingling in arms and legs with certain movements of the neck
- signs of a more systemic disease (such as rheumatoid arthritis or ankylosing spondylitis)

[a] Red flags are symptoms and/or signs for potentially serious specific conditions.
Signs compiled and slightly modified from Nachemson and Vingård, 2000; Binder, 2007; Rubinstein and Van Tulder, 2008; Nordin et al., 2009.

2.2.1. Inspection

Abnormal inspection signs such as muscle wasting, swelling, redness, scars in the neck are reported to have an interobserver reliability ranging from K = 0.32 to K = 0.81 [74]. Postural inspection to identify normal, excessive or diminished kyphosis of segments C7-Th2, Th3-5 and Th6-10 had an interobserver reliability ranging from K = 0.58 to K = 0.90 [21]. Postural observation of a forward head posture was not reliable [21].

2.2.2. Active and passive movements

Different aspects are assessed with the active and passive movements: pain reproduction, range of motion (ROM) and visual estimation of limitation of movements. Interobserver reliability for pain reproduction with active cervical movements ranged from K= 0.00 to K = 0.80 with a mean for all movements of K= 0.45 [21].

Quantification of passive and active ROM using a device has been reported as more reliable than visual estimation of ROM. However, others reported that active ROM visual estimated by clinicians was as reliable as using an external device with moderate interobserver reliability [74]. The most frequently used device is the cervical range of motion device (CROM). In previous studies the ICCs for interobserver reliability of active ROM ranged from 0.84 to 0.94 in patients with neck pain [37, 115].

Visual estimation of passive movements as limited or not limited in patients with neck pain had an interobserver reliability ranging from K = 0.05 to K = 0.85 in the principal directions [43, 80] and for all passive movements K = 0.54 [43]. Measurement of retraction of the head was less reliable than measurement of movements in de principal directions [40].

2.2.3. Intersegmental movements and pain provocation

Intersegmental cervical spine motion had an overall interobserver reliability ranging from K = 0.01 to K = 0.74 [21, 80]. Posterior to anterior pressure testing of C2 to T9 performed posteriorly over the spinous processes of the vertebrae to assess pain provocation had overall an interobserver reliability ranging from K = 0.00 to K = 0.90 [21].

2.2.4. Muscle strength and endurance tests

Muscle strength testing of the neck and upper extremity had slight to moderate reliability in patients with neck pain with or without radiculopathy [74]. Patients with CNP had slightly lower neck muscle strength compared with control subjects [74]. Cranio-cervical flexor strength (isometric maximal voluntary contraction (MVC)) and endurance at moderate (50% of MVC) and low (20% of MVC) loads were significantly lower in patients with neck pain [76].

2.2.5. Palpation of tender points

Palpation of tender points and trigger points in the neck region in a mixed group of patients with acute neck pain with or without arm pain and with CNP had an interobserver reliability ranging from K = 0.24 to K = 0.56 [101]. Palpation of paraspinal tender points is an appropriate screening test: the outcome of this test was consistent with the self-reported neck pain in questionnaires [88]. With clinical history and paraspinal tenderness palpation a reasonable inference can be drawn on the segmental location of the pain, but the actual structural source (e.g. muscle, or joint or intervertebral disc) cannot be determined [9]. Tender point distribution did not discriminate between patients with neck pain alone, neck pain and disc bulging shown on Magnetic Resonance Imaging (MRI) or neck pain with radiculopathy [47].

In patients with chronic pain tender point count is an indicator of emotional stress [114] and is positively associated with fatigue, depression and sleep problem scores [27]. State anxiety (anxiety that comes on with the starting of some undertaking) is a predictor of tender point count independently of pain severity and duration in CNP patients [31, 66].

2.2.6. Additional tests

The most widely used additional tests are the provocation tests for neck pain with radicular involvement. The (original) Spurling test, distraction test, upper limb tension test (ULTT), cervical rotation to the involved side less than 60˚ and Valsalva test are reliable tests with acceptable diagnostic accuracy [86, 87, 108].

2.3. Diagnostic triage

Diagnosis is based on clinical assessment and is the foundation of therapy. In probably 95% or more cases of neck pain one cannot diagnose definite pathology and neck pain is labeled as non-specific or mechanical neck pain [35]. Mechanical neck pain is characterized by: intermittent or constant pain in neck and shoulder region and/or between the shoulders, pain and other symptoms are influenced by posture and movements during physical activities and are changing in intensity during the day. Mechanical neck pain often develops spontaneously without specific etiology. It mostly presents itself in the early to middle years of adult life and the patient is not ill (general health not affected) [9, 35, 107].

Clinicians often fear that they will miss serious spine pathology in patients presenting themselves with neck pain. However, the risk is very low because on the one hand the incidence of specific neck disorders is very low and on the other hand serious spine pathology can be excluded with relatively simple screening methods [9, 35, 107]. For example in accident and emergency settings, only 3% of the patients with a clinical history of trauma and suspected injury had fractures of the cervical spine established with cervical radiography. So fractures of the cervical spine are actually uncommon [68]. Just as well the one year incidence of spinal

tumors is very low: between 0.1% and 0.7% depending the study population [91]. Arthritis of the zygapophyseal joints such as rheumatoid arthritis, ankylosing spondylitis, reactive arthritis, psoriatic arthritis and cristal arthropathies is very uncommon and is a rare cause of neck pain in patients without other manifestations of these disorders [9]. The largest group of specific neck disorder is neck pain with nerve root pain (radiculopathy) and neck pain with myelopathy due to cervical spondylosis. Although about 70% of the patients with neck pain in the general population do have brachialgia [97] the incidence of cervical radiculopathy is estimated to be 0.83 per thousand person-years [84]. It was suggested that about 5% of the patients with neck pain in the Dutch general practice do have radiculopathy [99].

Despite clear characteristics of mechanical neck pain in patients with chronic or repeated episodes of neck pain with high levels of stress and disability claims, clinicians may feel inclined to refer for secondary or tertiary diagnostic procedures [35]. A well-structured diagnostic triage is therefore helpful and gives the clinician the self-confidence to diagnose with little risk of error [9, 35, 87, 107].

The first step in the triage is to distinguish patients with neck pain of musculoskeletal origin from those with extrinsic neck pain (referred pain from other sites such as cardiac and diaphragmatic diseases). To rule out extrinsic neck pain the basic clinical history of key symptoms of these diseases is sufficient [8, 9]. If the history raises suspicion, further examination or diagnostic procedures are necessary.

The second step in the triage is to distinguish patients with 'simple' mechanical neck pain from those with specific neck pain with the help of red flags: symptoms or signs in clinical history and during physical examination that should raise the suspicion of specific neck disorders [9, 35, 87, 107]. An isolated red flag does not necessarily imply the presence of a serious specific neck disorder, however multiple positive red flags should raise clinical suspicion. Hence, it is of importance during the physical examination to confirm or potentially rule out suspicions raised during the history. To rule out a cervical fracture in patients with history of trauma and suspected injury a set of clinical symptoms and signs can be used: immediate onset of pain, head injury, loss of consciousness, midline cervical tenderness, impaired range of motion and neurological signs, [68]. Cervical radiculopathy can be the result of cervical degenerative changes and/or cervical disc protrusion/prolaps or other diseases. The pain characteristics, pattern of pain distribution and neurological signs suggest that cervical radiculopathy is most probably the case or not. When the provocative tests of the neck (Spurling, neck distraction and Valsalva) are positive, this might be indicative of a cervical radiculopathy [86]. To rule out cervical radiculopathy a negative upper limb tension test might be used [86]. Neck pain with myelopathy due to cervical spondylosis mostly develops insidiously. The clinical presentation is often with paresthesia and clumsiness of the hands and gait disturbance due to sensory ataxia of spastic paraparesis of the lower limbs [8]. When extrinsic neck pain and specific neck pain are ruled out the neck pain should be classified as non-specific.

Especially in sub-acute and chronic non-specific neck pain the third step of diagnostic triage is to assess orange flags as markers for psychopathology and to assess yellow, blue and black flags as markers for prolonged disability and failure to return to work [35, 53, 92, 107]. They offer the clinicians – especially in pain/occupational rehabilitation – a useful screening method when more in-depth analysis of psychopathology and/or psychosocial factors is warranted.

2.4. General practice

A total of 42.8 patients per 1000 registered persons in the Dutch general practice contact the general practitioner (GP) with complaints of the neck of which 19 patients with a new complaint or a new episode of a complaint of the neck [12]. The exact percentage of acute (<6 weeks), sub-acute (6–12 weeks) and chronic (>12 weeks) patients visiting their GP is unknown. Once neck pain has become chronic the minority (44%) of patients visits their GP on a yearly basis [10]. In patients with acute neck pain procedures most frequently applied by the GP in The Netherlands are physical examination (97%) and referral for further diagnostic investigation (X-ray 8%, blood test 1%, neurologist 1%) [103]. In patients with CNP GPs most frequently applied procedures in The Netherlands are physical examination (66%), referral for further diagnostic investigation (laboratory examinations (7%), X-ray (14%), medical specialist (14%) and other imaging techniques (CT, MRI, myelography, discography (2%) [10].

For neck pain most referrals to medical specialists by GPs are to neurologists, neurosurgeons, and orthopedic surgeons. Only a few are referred to rheumatologists and rehabilitation specialists [10].

2.5. Diagnostic imaging and invasive techniques

Patients with non-specific neck pain do not have a major structural disease. The clinical history and physical examination are more focused on excluding any specific disorder than on diagnostic imaging and other special diagnostic assessment tools [74]. Nevertheless plain radiographs and cervical MRI are relatively often performed. Reports about the degree of cervical lordosis or kyphosis and degenerative findings are often used by clinicians to 'explain' the symptoms of the patients. However, no significant difference is reported for example in cervical lordosis or kyphosis between patients with 'whiplash' exposure and healthy controls as documented by radiograph [65]. Moreover there is a similarly high prevalence of degenerative findings of the cervical spine in plain radiograph and MRI in asymptomatic individuals with a linear relationship between age and degenerative changes [61, 65]. Degenerative changes in the cervical spine are not a risk factor for neck pain and there is no evidence that they are strongly correlated with neck pain [74]. The Task Force on Neck Pain and its Associated Disorders reports also that in evaluating neck pain no evidence supports the use of cervical provocative discography, anesthetic facet or medical branch blocks [74].

3. THERAPY

3.1. Self-treatment

A large proportion of individuals with neck pain does not seek (para)medical care but copes without any (para)medical care. The most frequently used self-management strategies are analgetics, heat (infrared lamp, warm blanket/shower/bath, sauna, warm cloths, solarium), exercises (neck loosening exercises or auto manipulation, improving posture, fitness training), "trying another pillow", keeping the neck in rest or in immobilization (using a collar) [103].

3.2. General practitioner

GP's prescription in patients with acute neck pain are physiotherapy (51%), medication (42%) (mostly non-steroidal anti-inflammatory drugs 56%, or muscle relaxation medication 20%), instruction in home exercises (9%) and a wide range of advice [103]. Advice most frequently given by the GP is: to wait and see for an expected favorable natural course (23%), to improve posture and keep moving (22%) and to rest (18%) [103]. In patients with CNP therapeutic prescription by GPs are medication (79%) (mostly Paracetamol, Aspirin and non-steroidal anti-inflammatory drugs), physiotherapy (51%), heat application (20%), postural advice (18%) and rest (11%) [10].

3.3. Physiotherapy

In an inventory study of physiotherapy in patients with non-specific neck pain, shoulder pain and arm pain, the treatment (mean 11 sessions) was composed of exercise therapy (95%), massage (86%), electrotherapy (26%) and manual therapy techniques [50]. Goals in physiotherapy are to reduce pain, to improve strength or range of motion, to improve posture, to stretch muscles, to learn to relax, to improve functional movements [46]. In a randomized controlled trial in patients with non-specific neck pain most of the treatment time in physiotherapy was spent on exercise therapy, massage and instructions for home exercises [46].

A review of exercises for patients with mechanical neck pain reported that there is a role for exercises in the treatment of acute and chronic neck pain although the benefit of each type of exercise (endurance, strengthening, stretching, range of motion improvement, eye fixation exercises) is unclear [52]. Another review reported that supervised exercise interventions are more effective than no treatment, sham or alternative interventions; moreover, interventions focusing on regaining function as soon as possible are relatively more effective than interventions which are not [48]. In a review of massage intervention for mechanical neck pain no recommendations for or against massage were reported because of poor study quality and inconclusive results [33]. In a review of electrotherapy for mechanical neck pain the authors reported to have been unable to make any definite conclusions regarding the effectiveness [58].

In a randomized controlled trial usual physiotherapy produced marginally better treatment outcomes at 12 months than a brief physiotherapy intervention (one to three sessions) using cognitive behavioral principles to encourage self-management and to return to normal function [54]. Behavioral aspects in the treatment of patients with neck pain are gradually getting more attention in the physiotherapy setting in The Netherlands. Goals of this hands-off approach are: decrease in pain behavior and increase in "healthy behavior", improvement of function and no focus on pain reduction, patient is responsible for the treatment and has an active role and the physiotherapist acts as a coach [82]. In two randomized controlled trials in the Netherlands the effect of a behavioral graded activity (BGA) program was assessed. In patients with sub-acute non-specific neck pain there were no clinically relevant differences in improvement between a BGA program (mean visits 8.2) and manual therapy (mean visits 5.2) [82]. Moreover cost-effectiveness analyses showed that BGA is not cost-effective in comparison with manual therapy for recovery and gained quality-adjusted years of life [11]. In patients with chronic neck pain no significant differences between a BGA program (mean visits 6.6) and physiotherapy as usual (mean vistis 11.2) were observed [102].

3.4. Manual therapy

Manual therapy/medicine can be regarded as a post graduated specialization in diagnostic and conservative management of the musculoskeletal system. In The Netherlands manual therapy is mainly performed by physical therapists and partly by physicians. Both groups have had additional education and training, usually part time for 3 to 4 years. Treatment options that are often used are specific articular mobilizations and manipulations, instruction in functional home exercises, muscular (coordination, stabilization) techniques, clear explanation of symptoms and signs and education (ergonomic, ADL activities, changing life style, stimulating to be more active) [46, 49]. The majority of treatment time is spent on specific articular movement techniques, muscular techniques and instructions for home exercises [46]. Goals in manual therapy are: to restore range of motion, to stimulate natural recovery and adaptive processes related to functional movements, to reduce pain, to increase the patients level of activities and participation and to prevent recurrences [49, 82].

In a review, manual mobilizations and manipulations alone or with advice and home exercises were positively associated with better pain and functional outcomes in the short term (4 to 13 weeks) in patients with sub-acute or chronic neck pain [48]. However none of the aforementioned treatments is clearly superior to any other in the short or long term [48]. Another review reported that manual mobilizations and manipulations produced similar effects on pain, functional status and patient satisfaction in the short or long term [38].

In a recent randomized controlled trial the effectiveness of spinal manipulation therapy (SMT), medication, and supervised instruction in home exercise with advice (HEA) in patients with acute and sub-acute neck pain was compared [15]. The intervention period was 12 weeks. For

pain and most of the secondary outcomes SMT (mean visits 15.3) and HEA (mean visits 2.0) were similarly effective and both were more effective than medication (mean visits 4.8) at most follow-up time points [15].

In a randomized controlled trial in The Netherlands the effectiveness of manual therapy, physical therapy and continued care by the GP (analgesics, counseling and education) was compared [46]. Manual therapy (median visits 6.0) had significantly better results than physiotherapy (median visits 9.0) and continued care by the GP (median visits 2.0). Patients with more severe complaints and a high physical dysfunction severity appear to benefit most from manual therapy [46]. The direct and indirect costs of manual therapy are approximately one third of the costs of physical therapy and care by the GP [56].

3.5. Multidisciplinary pain rehabilitation

Patients are often referred to a rehabilitation center because of failure of the most standard biomedical therapy with requests for admission to multidisciplinary pain rehabilitation according to a biopsychosocial pain model. Thus patients treated at rehabilitation centers are the more chronic and complex cases. The multidisciplinary pain rehabilitation includes multiple treatment components that require close coordination between physicians, psychologists, social workers, vocational experts, physiotherapists and occupational therapists. Common components of a multidisciplinary rehabilitation program are cognitive restructuring, behavioral therapy, physical training, ergonomic training, patient education and vocational rehabilitation [96]. Goals are to restore physical and psychosocial activities, to increase activity levels at home and in day to day functioning and to facilitate return to work [17, 96]. No evidence was found to support the effectiveness of multidisciplinary outpatient rehabilitation at private clinics or hospital inpatient rehabilitation in chronic whiplash patients [20]. A Cochrane review reported limited evidence for the effectiveness of multidisciplinary rehabilitation for neck and shoulder pain [51]. The Cochrane Back Group concluded that there was an urgent need for high quality trials in this field [51].

4. MEASURING NECK PAIN RELATED DISABILITY

The International Classification of Functioning Disability and Health (ICF; Figure 1.1) defines disability as an umbrella term for impairments, activity limitations and participation restrictions [112]. Patients with neck pain may be impaired in body functions and structures, limited in performing activities and restricted in participation [95, 112]. Impairments in body functions and structure cannot fully explain neck pain related disability. Therefore the ICF model is a biopsychosocial model which also incorporates activity limitations, participation restriction and environmental and personal factors that influence one's perception of functioning and disability [95, 112].

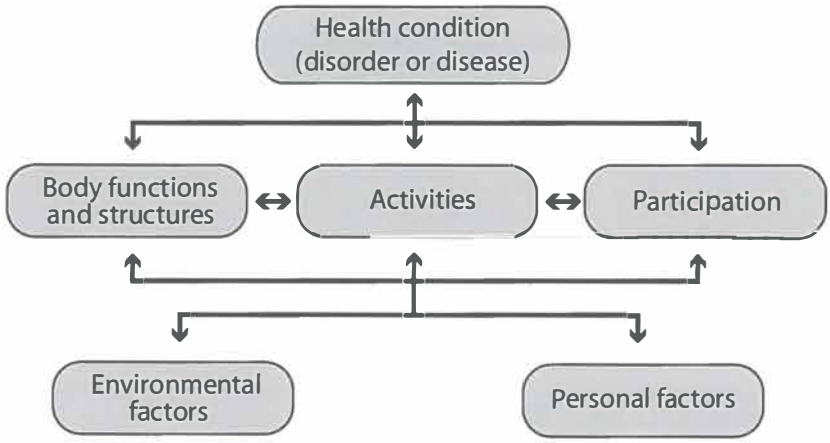


Figure 1.1 Model of the International Classification of Functioning, Disability and Health [112].

Measuring disability is important in rehabilitation research and practice because it can be used for quantifying current disability status, predicting future disability and evaluating treatment programs. Disability can be measured using different methods such as questionnaires, functional testing (for instance by a functional capacity evaluation (FCE)), clinical assessment (based on clinical history and physical examination) or a combination of these methods [16].

Comparative research between a patient's self-report of perceived disability, expert based assessment of the disability by the physician and performance based assessment of the disability with functional testing by an experienced physiotherapist has shown little similarity and low correlation in patients with chronic low back pain [16]. Self-reported disability was considerably higher than the expert based disability estimated by the physician and the performance based disability assessed by the physiotherapist experienced in functional capacity evaluation. Each method of measuring disability deals with a different perspective and probably with a different aspect of disability.

An instrument aiming to measure the impact of the cervical pain and dysfunction on the performance of common daily activities and to measure change in health status should have good clinimetric qualities; reproducibility, validity and responsiveness to change [79].

Reproducibility is the extent to which the same results are obtained on repeated tests when no real change in construct has occurred [29, 79]. Reproducibility may be influenced by random variation of measurement results and within patient variations [29, 83]. Both sources of variation may lead to "instability" of "fluctuations" in health status in the absence of treatment; the "natural variations" or the "background noise" [29, 83].

Validity refers to the extent to which the instrument measures what is intended to measure [79]. Various forms of validity are commonly described: content-, criterion- and construct validity [79, 83]. A scale is considered to have content validity when the items reflect all aspects of the construct being measured. Content validity cannot be tested but is inferred from expert judgments. Criterion validity is a degree to which the scores of an instrument are an adequate reflexion of a 'gold standard' [69]. Construct validity refers to the extent to which the scores of an instrument correlates as hypothesized with other assessment tools (for example neck pain scale versus a quality-of-life questionnaire) [69, 79, 83]. Responsiveness is the ability of an instrument to detect change over time in the construct to be measured [69].

5. MEASUREMENT INSTRUMENTS

5.1. Self-reported disability

In rehabilitation medicine, self-reported disability in patients with neck pain is often measured by means of questionnaires. The most frequently used neck disability questionnaires are the Neck Pain and Disability Scale (NPAD) and the Neck Disability Index (NDI) [100, 111]. The most frequently used general health questionnaire is the Short Form-36 Health Survey (SF-36) [110].

5.1.1. NPAD

The NPAD consists of 20 items [111]. Each item has a Visual Analogue Scale (VAS) of 100 mm with numeric anchors at 0, 1, 2, 3, 4 and 5 (each 20 mm apart). Item scores range from 0 (no pain or activity limitation) to 5 (as much pain as possible or maximal limitation). The total NPAD score ranges from 0 to 100 points. Higher scores indicate greater disability [111]. The NPAD has shown to be a reliable and valid measure of disability in different languages [6, 14, 23, 60, 70, 71, 113].The NPAD has not been formally translated into Dutch and consequently clinimetric qualities of the Dutch Language Version (NPAD-DLV) are unknown.

5.1.2. NDI

The NDI consists of 10 items [100]. Each item has 6 different assertions expressing progressive levels of pain or limitation in activities. Item scores range from 0 (no pain or limitation) to 5 (as much pain as possible or maximal limitation. The total NDI score ranges from 0 to 50 points. Higher scores indicate greater disability [100]. The NDI has shown to be a reliable and valid measure of disability in different languages [5, 23, 57, 71, 104, 113]. The clinimetric qualities of the NDI-DLV in patients with CNP are unknown.

In a systematic review evaluating 8 different neck questionnaires the NDI seems to have the most adequate measurement properties except for reliability [90]. The disadvantage of the NDI

is that the 6 response assertions of each of the10 items are wordy, inconsistent and confusing [30, 98]. In 6 of 10 items more than one concept is incorporated within the response assertions (e.g. pain and need of help in the 'personal care' item and lifting ability and pain level in the 'lifting' item) [30, 98]. The advantage of the NPAD over the NDI may be the simple wording and the unitary structure of the questions; moreover the questionnaire is easy to complete with its visual analog scale structure [14, 23, 79].

5.1.3. SF-36

The SF-36 is a general health questionnaire consisting of 36 items [110]. It assesses health over the past 4 weeks in 8 domains: physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health. Scores for each domain range from 0 to 100, with higher scores indicating a better health. The physical component summary (PCS) and mental component summary (MCS) are constructed by using weighted aggregates of all 8 domains of the SF-36 [109]. Higher PCS or MCS scores also imply better health. The SF-36-DLV has shown to be a reliable and valid instrument in community populations and in patients with chronic disease [1]. The construct validity of the SF-36 has not been studied in patients with non-specific CNP in an outpatient tertiary rehabilitation setting.

5.2. Clinical examination

The clinical history provides probably the most relevant information for diagnostic and treatment goals. Nevertheless a physical examination based on the assessment of the mobility of the cervical spine, pain provocation with cervical movements [46] and behavioral signs [59, 72] provides relevant information with regard to aspects of the patient's functional status. Therefore it is important to know the reliability of these assessments.

5.2.1. Physical dysfunction severity (PDS)

As measure for actual physical impairment of the cervical spine, the physical dysfunction severity (PDS) scale was used for evaluation of treatment efficacy in primary care [46, 55]. The PDS was based on severity of neck pain and limitations in spinal mobility during physical examination [46]. The interobserver reliability and construct validity of the PDS has not been investigated.

5.2.2. Cervical non organic signs (cNOS)

A standardized set of 'non organic signs' (NOS) or behavioral signs was developed for patients with back pain [106] and neck pain [92] to screen them for yellow flags in the physical examination [53, 72]. Multiple behavioral signs suggest that the patient's complaints cannot be explained only from a biomedical perspective only and that psychosocial factors may need closer consideration [63, 106]. In back pain NOS were associated with poor nonsurgical treatment

outcome, greater levels of pain and decreased functional performance [36, 75]. Cervical NOS (cNOS) were associated with a higher risk for prolonged disability [59]. Prevalence of positive signs of cNOS as well as interobserver reliability for total scores of cNOS were not reported. Moreover the construct validity of cNOS is unknown. Because of reasons described in chapter 6 of this thesis we changed the cNOS tests in the categories simulation and distraction into the modified cNOS (mcNOS).

5.3. Functional capacity evaluation (FCE)

An FCE is a standardized battery of tests to evaluate a patient's capacity to perform work-related activities. Because in the ICF model biopsychosocial factors such as personal and contextual factors are also taken into consideration, the use of a more comprising definition was suggested: 'An FCE is an evaluation of capacity of activities that is used to make recommendations for participation in work, while considering the person's body functions and structures, environmental factors, personal factors and health status' [93]. FCEs are used in work rehabilitation programs for measurement of disability return to work recommendations and in medico-legal issues. Therefore an FCE specific for patients with work-related neck disorders is needed. However, no validated neck specific FCE has been described in peer reviewed literature.

6. AIMS AND RESEARCH QUESTIONS ADDRESSED IN THIS THESIS

In this thesis some of the clinimetric properties of the instruments mentioned will be investigated in patients with CNP in an outpatient tertiary rehabilitation setting. The following research questions will be addressed in this thesis:

NPAD/NDI

- What is the test-retest reproducibility of the Dutch language version of the NPAD and NDI?
- What is the content validity, internal consistency and construct validity of the Dutch language versions of these questionnaires?
- What are relevant changes in total scores on the NPAD and NDI after completion of a multidisciplinary pain rehabilitation program and which questionnaire is most responsive to change?

SF-36

- What is the construct validity of the SF-36?

PDS

- What is the interobserver reliability of the PDS?
- What is the construct validity of the PDS?

mcNOS

- What is the interobserver reliability of the mcNOS?
- What is the construct validity of the mcNOS?

FCE

- Can a content valid FCE be developed for patients with work-related neck pain?

7. OUTLINE OF THIS THESIS

In chapter 2 a study is presented in which a translation and cross-cultural adaptation of the NPAD is performed and the test-retest reproducibility of the NPAD and NDI is presented. The study focuses on 2 aspects of reproducibility: test-retest reliability and agreement.

In chapter 3 a study is presented in which the validity of the NPAD and NDI is tested based on 27 a priori hypotheses. The study focuses on content validity, internal consistency and construct validity.

In chapter 4 a study is presented in which relevant changes in total scores of NPAD and NDI are assessed with minimal detectable change and minimal important change after completion of a multidisciplinary rehabilitation program. Additionally, a comparison is made between the two questionnaires regarding responsiveness to change in rehabilitation treatment.

In chapter 5 a study is presented in which the construct validity of the SF-36 is tested based on 16 a priori hypotheses. The study presents the comparison of the SF-36 domain scores of patients with non-specific CNP with general population norm values. Additionally the study presents the comparison of SF-36 physical and mental component summary scores with values of patients in a primary care setting.

In chapter 6 a study is presented which focuses on the interobserver reliability of the PDS as measure of actual physical impairment of the cervical spine and mcNOS as measure for behavioral signs. Additionally, the study focuses on the construct validity of the PDS and mcNOS.

In chapter 7 a study is presented in which an FCE was developed which has content validity for determining functional capacity in patients with work-related neck disorders.

Finally, chapter 8 provides a general discussion regarding the clinimetric properties of the assessments. The clinical relevance of the outcomes of the studies is discussed and directions for further research are presented.

REFERENCES

1. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, Sprangers MA, te Velde A, Verrips E (1998) Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol 51:1055–1068
2. Andersson HI, Ejlertsson G, Leden I, Schersten B (1999) Musculoskeletal chronic pain in general practice. Studies of health care utilisation in comparison with pain prevalence. Scand J Prim Health Care 17:87–92
3. Ariens GA, van Mechelen W, Bongers PM, Bouter LM, van der Wal G (2000) Physical risk factors for neck pain. Scand J Work Environ Health 26:7–19
4. Ariens GA, van Mechelen W, Bongers PM, Bouter LM, van der Wal G (2001) Psychosocial risk factors for neck pain: a systematic review. Am J Ind Med 39:180–193
5. Aslan E, Karaduman A, Yakut Y, Aras B, Simsek IE, Yagly N (2008) The cultural adaptation, reliability and validity of neck disability index in patients with neck pain: a Turkish version study. Spine (Phila Pa 1976) 33:E362–E365
6. Bicer A, Yazici A, Camdeviren H, Erdogan C (2004) Assessment of pain and disability in patients with chronic neck pain: reliability and construct validity of the Turkish version of the neck pain and disability scale. Disabil Rehabil 26:959–962
7. Binder A (2006) Neck pain. Clin Evid (15):1654–1675
8. Binder A (2007) The diagnosis and treatment of nonspecific neck pain and whiplash. Eura Medicophys 43:79–89
9. Bogduk N (1999) The neck. Baillieres Clin Rheumatol 13:261–285
10. Borghouts JA, Koes BW, Vondeling H, Bouter LM (1999) Cost-of-illness of neck pain in The Netherlands in 1996. Pain 80:629–636
11. Bosmans JE, Pool JJ, de Vet HC, van Tulder MW, Ostelo RW (2011) Is behavioral graded activity cost-effective in comparison with manual therapy for patients with subacute neck pain? An economic evaluation alongside a randomized clinical trial. Spine (Phila Pa 1976) 36:E1179–86
12. Bot SD, van der Waal JM, Terwee CB, van der Windt DA, Schellevis FG, Bouter LM, Dekker J (2005) Incidence and prevalence of complaints of the neck and upper extremity in general practice. Ann Rheum Dis 64:118–123
13. Bovim G, Schrader H, Sand T (1994) Neck pain in the general population. Spine (Phila Pa 1976) 19:1307–1309
14. Bremerich FH, Grob D, Dvorak J, Mannion AF (2008) The Neck Pain and Disability Scale: cross-cultural adaptation into German and evaluation of its psychometric properties in chronic neck pain and C1-2 fusion patients. Spine 33:1018–1027
15. Bronfort G, Evans R, Anderson AV, Svendsen KH, Bracha Y, Grimm RH (2012) Spinal manipulation, medication, or home exercise with advice for acute and subacute neck pain: a randomized trial. Ann Intern Med 156:1–10

16. Brouwer S, Dijkstra PU, Stewart RE, Goeken LN, Groothoff JW, Geertzen JH (2005) Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. Disabil Rehabil 27:999–1005

17. Buchner M, Zahlten-Hinguranage A, Schiltenwolf M, Neubauer E (2006) Therapy outcome after multidisciplinary treatment for chronic neck and chronic low back pain: a prospective clinical study in 365 patients. Scand J Rheumatol 35:363–367

18. Carroll LJ, Hogg-Johnson S, van der Velde G, Haldeman S, Holm LW, Carragee EJ, Hurwitz EL, Cote P, Nordin M, Peloso PM, Guzman J, Cassidy JD, Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders (2008) Course and prognostic factors for neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Spine (Phila Pa 1976) 33:S75–82

19. Carroll LJ, Hogg-Johnson S, van der Velde G, Haldeman S, Holm LW, Carragee EJ, Hurwitz EL, Cote P, Nordin M, Peloso PM, Guzman J, Cassidy JD (2009) Course and prognostic factors for neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. J Manipulative Physiol Ther 32:S87–96

20. Cassidy JD, Carroll LJ, Cote P, Frank J (2007) Does multidisciplinary rehabilitation benefit whiplash recovery?: results of a population-based incidence cohort study. Spine (Phila Pa 1976) 32:126–131

21. Cleland JA, Childs JD, Fritz JM, Whitman JM (2006) Interrater reliability of the history and physical examination in patients with mechanical neck pain. Arch Phys Med Rehabil 87:1388–1395

22. Cloward RB (1959) Cervical diskography. A contribution to the etiology and mechanism of neck, shoulder and arm pain. Ann Surg 150:1052–1064

23. Cook C, Richardson JK, Braga L, Menezes A, Soler X, Kume P, Zaninelli M, Socolows F, Pietrobon R (2006) Cross-cultural adaptation and validation of the Brazilian Portuguese version of the Neck Disability Index and Neck Pain and Disability Scale. Spine 31:1621–1627

24. Cote P, Cassidy JD, Carroll L (1998) The Saskatchewan Health and Back Pain Survey. The prevalence of neck pain and related disability in Saskatchewan adults. Spine (Phila Pa 1976) 23:1689–1698

25. Cote P, Cassidy JD, Carroll L (2001) The treatment of neck and low back pain: who seeks care? who goes where? Med Care 39:956–967

26. Cote P, Cassidy JD, Carroll LJ, Kristman V (2004) The annual incidence and course of neck pain in the general population: a population-based cohort study. Pain 112:267–273

27. Croft P, Schollum J, Silman A (1994) Population study of tender point counts and pain as evidence of fibromyalgia. BMJ 309:696–699

28. Croft PR, Lewis M, Papageorgiou AC, Thomas E, Jayson MI, Macfarlane GJ, Silman AJ (2001) Risk factors for neck pain: a longitudinal study in the general population. Pain 93:317–325

29. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ (2001) Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 17:479–487

30. Dijkstra PU, van der Schans CP (2002) Ingezonden brief. Reactie op 'Neck Disability Index Dutch Version bij chronische "whiplash"-patiënten: onderzoek naar de betrouwbaarheid'. Ned T Fysiother 112 (6):163

31. Dyrehag LE, Widerstrom-Noga EG, Carlsson SG, Kaberger K, Hedner N, Mannheimer C, Andersson SA (1998) Relations between self-rated musculoskeletal symptoms and signs and psychological distress in chronic neck and shoulder pain. Scand J Rehabil Med 30:235–242

32. En MC, Clair DA, Edmondston SJ (2009) Validity of the Neck Disability Index and Neck Pain and Disability Scale for measuring disability associated with chronic, non-traumatic neck pain. Man Ther 14:433–438

33. Ezzo J, Haraldsson BG, Gross AR, Myers CD, Morien A, Goldsmith CH, Bronfort G, Peloso PM, Cervical Overview Group (2007) Massage for mechanical neck disorders: a systematic review. Spine (Phila Pa 1976) 32:353–362

34. Fejer R, Hartvigsen J (2008) Neck pain and disability due to neck pain: what is the relation? Eur Spine J 17:80–88

35. Ferrari R, Russell AS (2003) Regional musculoskeletal conditions: neck pain. Best Pract Res Clin Rheumatol 17:57–70

36. Fishbain DA, Cole B, Cutler RB, Lewis J, Rosomoff HL, Rosomoff RS (2003) A structured evidence-based review on the meaning of nonorganic physical signs: Waddell signs. Pain Med 4:141–181

37. Fletcher JP, Bandy WD (2008) Intrarater reliability of CROM measurement of cervical spine active range of motion in persons with and without neck pain. J Orthop Sports Phys Ther 38:640–645

38. Gross A, Miller J, D'Sylva J, Burnie SJ, Goldsmith CH, Graham N, Haines T, Bronfort G, Hoving JL, COG (2010) Manipulation or mobilisation for neck pain: a Cochrane Review. Man Ther 15:315–333

39. Haldeman S, Carroll L, Cassidy JD, Schubert J, Nygren A (2008) The Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders: executive summary. Spine 33:S5–S7

40. Hanten WP, Olson SL, Russell JL, Lucio RM, Campbell AH (2000) Total head excursion and resting head posture: normal and patient comparisons. Arch Phys Med Rehabil 81:62–66

41. Hill JC, Lewis M, Sim J, Hay EM, Dziedzic K (2007) Predictors of poor outcome in patients with neck pain treated by physical therapy. Clin J Pain 23:683–690

42. Hogg-Johnson S, van der Velde G, Carroll LJ, Holm LW, Cassidy JD, Guzman J, Cote P, Haldeman S, Ammendolia C, Carragee E, Hurwitz E, Nordin M, Peloso P, Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders (2008) The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Spine (Phila Pa 1976) 33:S39–51

43. Hoppenbrouwers M, Eckhardt MM, Verkerk K, Verhagen A (2006) Reproducibility of the measurement of active and passive cervical range of motion. J Manipulative Physiol Ther 29:363–367

44. Hoving JL, O'Leary EF, Niere KR, Green S, Buchbinder R (2003) Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. Pain 102:273–281

45. Hoving JL, de Vet HC, Twisk JW, Deville WL, van der Windt D, Koes BW, Bouter LM (2004) Prognostic factors for neck pain in general practice. Pain 110:639–645

46. Hoving JL, de Vet HC, Koes BW, Mameren H, Deville WL, van der Windt DA, Assendelft WJ, Pool JJ, Scholten RJ, Korthals-de Bos IB, Bouter LM (2006) Manual therapy, physical therapy, or continued care by the general practitioner for patients with neck pain: long-term results from a pragmatic randomized clinical trial. Clin J Pain 22:370–377

47. Hsueh TC, Yu S, Kuan TS, Hong CZ (1998) Association of active myofascial trigger points and cervical disc lesions. J Formos Med Assoc 97:174–180

48. Hurwitz EL, Carragee EJ, van der Velde G, Carroll LJ, Nordin M, Guzman J, Peloso PM, Holm LW, Cote P, Hogg-Johnson S, Cassidy JD, Haldeman S, Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders (2008) Treatment of neck pain: noninvasive interventions: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Spine (Phila Pa 1976) 33:S123–52

49. Jorritsma W, Brouwer R, Molenaar J, Jansen MPL (2002) Beroepsprofiel en eindtermen van de (ortho) manuele geneeskunde.

50. Karels C (2006) Behandeling en prognose van arm, nek en schouderklachten in de fysiotherapiepraktijk. Congres: De cervicale wervelkolom nader beschouwd. Rotterdam.

51. Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, Koes B (2003) Multidisciplinary biopsychosocial rehabilitation for neck and shoulder pain among working age adults. Cochrane Database Syst Rev (2):CD002194

52. Kay TM, Gross A, Goldsmith C, Santaguida PL, Hoving J, Bronfort G, Cervical Overview Group (2005) Exercises for mechanical neck disorders. Cochrane Database Syst Rev (3):CD004250

53. Kendall N, Linton S, Main C (1997) Guide to Assessing Psychosocial Yellow Flags in Acute Low Back Pain: Risk Factors for Long-Term Disability and Work Loss

54. Klaber Moffett JA, Jackson DA, Richmond S, Hahn S, Coulton S, Farrin A, Manca A, Torgerson DJ (2005) Randomised trial of a brief physiotherapy intervention compared with usual physiotherapy for neck pain patients: outcomes and patients' preference. BMJ 330:75

55. Koes BW, Bouter LM, van Mameren H, Essers AH, Verstegen GM, Hofhuizen DM, Houben JP, Knipschild PG (1992) A blinded randomized clinical trial of manual therapy and physiotherapy for chronic back and neck complaints: physical outcome measures. J Manipulative Physiol Ther 15:16–23

56. Korthals-de Bos IB, Hoving JL, van Tulder MW, Rutten-van Molken MP, Ader HJ, de Vet HC, Koes BW, Vondeling H, Bouter LM (2003) Cost effectiveness of physiotherapy, manual therapy, and general practitioner care for neck pain: economic evaluation alongside a randomised controlled trial. BMJ 326:911

57. Kose G, Hepguler S, Atamaz F, Oder G (2007) A comparison of four disability scales for Turkish patients with neck pain. J Rehabil Med 39:358–362

58. Kroeling P, Gross AR, Goldsmith CH, Cervical Overview Group (2005) A Cochrane review of electrotherapy for mechanical neck disorders. Spine (Phila Pa 1976) 30:E641–8

59. Landers MR, Creger RV, Baker CV, Stutelberg KS (2008) The use of fear-avoidance beliefs and nonorganic signs in predicting prolonged disability in patients with neck pain. Man Ther 13:239–248

60. Lee H, Nicholson LL, Adams RD, Maher CG, Halaki M, Bae SS (2006) Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. Spine 31:1841–1845

61. Lehto IJ, Tertti MO, Komu ME, Paajanen HE, Tuominen J, Kormano MJ (1994) Age-related MRI changes at 0.1 T in cervical discs in asymptomatic subjects. Neuroradiology 36:49–53

62. Linton SL (2000) Psychological risk factors for neck and back pain. In: Nachemson AL, Jonsson E (eds) Neck and Back Pain. Lippincott Williams & Wilkens, Philadelphia, pp 57–78

63. Main CJ, Waddell G (1998) Behavioral responses to examination. A reappraisal of the interpretation of "nonorganic signs". Spine (Phila Pa 1976) 23:2367–2371

64. Main CJ, Burton AK (2000) Economic and occupational influences on pain and disability. In: Main CJ, Spanswick CC (eds) Pain Management: An Interdisciplinary Approach. Churchill Livingstone, Edinburgh, Scotland, pp 63–87

65. Matsumoto M, Fujimura Y, Suzuki N, Nishi Y, Nakamura M, Yabe Y, Shiga H (1998) MRI of cervical intervertebral discs in asymptomatic subjects. J Bone Joint Surg Br 80:19–24

66. McBeth J, Macfarlane GJ, Benjamin S, Morris S, Silman AJ (1999) The association between tender points, psychological distress, and adverse childhood experiences: a community-based study. Arthritis Rheum 42:1397–1404

67. McKenzie R, May S (2006) The Cervical & Thoracic Spine: Mechanical Diagnosis & Therapy. Spinal Publications New Zealand Ltd, Raumati Beach New Zealand

68. McNamara RM, O'Brien MC, Davidheiser S (1988) Post-traumatic neck pain: a prospective and follow-up study. Ann Emerg Med 17:906–911

69. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. BMC Med Res Methodol 10:22

70. Monticone M, Baiardi P, Nido N, Righini C, Tomba A, Giovanazzi E (2008) Development of the Italian version of the Neck Pain and Disability Scale, NPDS-I: cross-cultural adaptation, reliability, and validity. Spine 33:E429–E434

71. Mousavi SJ, Parnianpour M, Montazeri A, Mehdian H, Karimi A, Abedi M, Ashtiani AA, Mobini B, Hadian MR (2007) Translation and validation study of the Iranian versions of the Neck Disability Index and the Neck Pain and Disability Scale. Spine 32:E825–E831

72. Nachemson A, Vingard E (2000) Assessment of patients with neck and back pain: A best-evidence synthesis. In: Nachemson AL, Jonsson E (eds) Neck and Back Pain. Lippincott Williams & Wilkens, Philadelphia, pp 189–236

73. Nicholas MK, Linton SJ, Watson PJ, Main CJ, "Decade of the Flags" Working Group (2011) Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Phys Ther 91:737–753

74. Nordin M, Carragee EJ, Hogg-Johnson S, Weiner SS, Hurwitz EL, Peloso PM, Guzman J, van der Velde G, Carroll LJ, Holm LW, Cote P, Cassidy JD, Haldeman S (2009) Assessment of neck pain and its associated disorders: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. J Manipulative Physiol Ther 32:S117–40

75. Oesch P, Meyer K, Jansen B, Mowinckel P, Bachmann S, Hagen KB (2012) What is the role of "nonorganic somatic components" in functional capacity evaluations in patients with chronic nonspecific low back pain undergoing fitness for work evaluation? Spine (Phila Pa 1976) 37:E243–50

76. O'Leary S, Jull G, Kim M, Vicenzino B (2007) Cranio-cervical flexor muscle impairment at maximal, moderate, and low loads is a feature of neck pain. Man Ther 12:34–39

77. Palmer KT, Walker-Bone K, Griffin MJ, Syddall H, Pannett B, Coggon D, Cooper C (2001) Prevalence and occupational associations of neck pain in the British population. Scand J Work Environ Health 27:49–56

78. Picavet HS, Schouten JS (2003) Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC(3)-study. Pain 102:167–178

79. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine 27:515–522

80. Pool JJ, Hoving JL, de Vet HC, van Mameren H, Bouter LM (2004) The interexaminer reproducibility of physical examination of the cervical spine. J Manipulative Physiol Ther 27:84–90

81. Pool JJ, Ostelo RW, Knol D, Bouter LM, de Vet HC (2010) Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain? Man Ther 15:111–116

82. Pool JJ, Ostelo RW, Knol DL, Vlaeyen JW, Bouter LM, de Vet HC (2010) Is a behavioral graded activity program more effective than manual therapy in patients with subacute neck pain? Results of a randomized clinical trial. Spine (Phila Pa 1976) 35:1017–1024

83. Portney LG, Watkins MP (2000) Foundations of Clinical Research 2nd Edition Prentice Hall Health. , Upper Saddle River, New Yersey 07458

84. Radhakrishnan K, Litchy WJ, O'Fallon WM, Kurland LT (1994) Epidemiology of cervical radiculopathy. A population-based study from Rochester, Minnesota, 1976 through 1990. Brain 117 (Pt 2):325–335

85. Reitsma B, Meijler WJ (1997) Pain and patienthood. Clin J Pain 13:9–21

86. Rubinstein SM, Pool JJ, van Tulder MW, Riphagen II, de Vet HC (2007) A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. Eur Spine J 16:307–319

87. Rubinstein SM, van Tulder M (2008) A best-evidence review of diagnostic procedures for neck and low-back pain. Best Pract Res Clin Rheumatol 22:471–482

88. Sandmark H, Nisell R (1995) Validity of five common manual neck pain provoking tests. Scand J Rehabil Med 27:131–136

89. Schellingerhout JM, Heymans MW, Verhagen AP, Lewis M, de Vet HC, Koes BW (2010) Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. Spine (Phila Pa 1976) 35:E827–35

90. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB (2012) Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. Qual Life Res 21:659–670

91. Slipman CW, Patel RK, Botwin K, Huston C, Zhang L, Lenrow D, Garvan C (2003) Epidemiology of spine tumors presenting to musculoskeletal physiatrists. Arch Phys Med Rehabil 84:492–495

92. Sobel JB, Sollenberger P, Robinson R, Polatin PB, Gatchel RJ (2000) Cervical nonorganic signs: a new clinical tool to assess abnormal illness behavior in neck pain patients: a pilot study. Arch Phys Med Rehabil 81:170–175

93. Soer R, van der Schans CP, Groothoff JW, Geertzen JH, Reneman MF (2008) Towards consensus in operational definitions in functional capacity evaluation: a Delphi Survey. J Occup Rehabil 18:389–400

94. Stock SR (1991) Workplace ergonomic factors and the development of musculoskeletal disorders of the neck and upper limbs: a meta-analysis. Am J Ind Med 19:87–107

95. Stucki G (2005) International Classification of Functioning, Disability, and Health (ICF): a promising framework and classification for rehabilitation medicine. Am J Phys Med Rehabil 84:733–740

96. Turk DC, Wilson HD, Cahana A (2011) Treatment of chronic non-cancer pain. Lancet 377:2226–2235

97. Valkenburg HA (1980) Reumatische klachten en aandoeningen in een open Nederlandse bevolkingsgroep. Nek en lage rugklachten (deel 1). Reuma Wereldwijd 1:1–6

98. van der Velde G, Beaton D, Hogg-Johnston S, Hurwitz E, Tennant A (2009) Rasch analysis provides new insights into the measurement properties of the neck disability index. Arthritis Rheum 61:544–551

99. Verhagen AP, van Ruiven S, Thomas S (2009) Diagnosis and treatment of neck pain. Ned Tijdschr Geneeskd 153:558–561

100. Vernon H, Mior S (1991) The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 14:409–415

101. Viikari-Juntura E (1987) Interexaminer reliability of observations in physical examinations of the neck. Phys Ther 67:1526–1532

102. Vonk F, Verhagen AP, Twisk JW, Koke AJ, Luiten MW, Koes BW (2009) Effectiveness of a behaviour graded activity program versus conventional exercise for chronic neck pain patients. Eur J Pain 13:533–541

103. Vos C, Verhagen A, Passchier J, Koes B (2007) Management of acute neck pain in general practice: a prospective study. Br J Gen Pract 57:23–28

104. Vos CJ, Verhagen AP, Koes BW (2006) Reliability and responsiveness of the Dutch version of the Neck Disability Index in patients with acute neck pain in general practice. Eur Spine J 15:1729–1736

105. Vos CJ, Verhagen AP, Passchier J, Koes BW (2008) Clinical course and prognostic factors in acute neck pain: an inception cohort study in general practice. Pain Med 9:572–580

106. Waddell G, McCulloch JA, Kummel E, Venner RM (1980) Nonorganic physical signs in low-back pain. Spine (Phila Pa 1976) 5:117–125

107. Waddell G (1998) The Back Pain Revolution. Churchill Livingstone, Edinburgh

108. Wainner RS, Fritz JM, Irrgang JJ, Boninger ML, Delitto A, Allison S (2003) Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. Spine (Phila Pa 1976) 28:52–62

109. Ware JE, Kosinski M, Keller SD (1994) SF-36 physical and mental health summary scales: a user's manual. The Health Institute, New England Medical Center, Boston

110. Ware JE, Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473–483

111. Wheeler AH, Goolkasian P, Baird AC, Darden BV (1999) Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. Spine 24:1290–1294

112. WHO (2001) ICF: International Classification of Functioning, Disability, and Health. World Health Organization, Geneva

113. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel M (2002) French translation and validation of 3 functional disability scales for neck pain. Arch Phys Med Rehabil 83:376–382

114. Wolfe F (1997) The relation between tender points and fibromyalgia symptom variables: evidence that fibromyalgia is not a discrete disorder in the clinic. Ann Rheum Dis 56:268–271

115. Youdas JW, Carey JR, Garrett TR (1991) Reliability of measurements of cervical spine range of motion--comparison of three methods. Phys Ther 71:98–104; discussion 105–6

Chapter 2

Neck Pain and Disability Scale and the Neck Disability Index: reproducibility of the Dutch language versions

Wim Jorritsma
Grietje E. de Vries
Jan H. B. Geertzen
Pieter U. Dijkstra
Michiel F. Reneman

Eur Spine J (2010) 19:1695–1701

ABSTRACT

The first aim of this study was to translate the Neck Pain and Disability Scale (NPAD) from English into Dutch producing the NPAD–Dutch Language Version (DLV). The second aim was to analyze test–retest reliability and agreement of the NPAD–DLV and the Neck Disability Index (NDI)–DLV. The NPAD was translated according to established guidelines. Thirty-four patients (mean age 37.5 years, 68% female) with chronic neck pain (CNP), within an outpatient rehabilitation setting, participated in this study. The NPAD–DLV and the NDI–DLV were filled out twice with a mean test–retest interval of 18 days. The intraclass correlation coefficient of the NPAD–DLV was 0.76 (95% confidence interval (CI) 0.57–0.87) and of the NDI–DLV 0.84 (95% CI 0.69–0.92). The limits of agreement of the NPAD–DLV and the NDI–DLV were, respectively, ±20.9 (scale 0–100) and ± 6.5 (scale 0–50). The reliability of the NPAD–DLV and the NDI–DLV was acceptable for patients with CNP. The variation ('instability') in the NPAD–DLV total scores was relatively large and larger than the variation of the NDI–DLV.

INTRODUCTION

Neck pain is a common musculoskeletal complaint in western societies [15]. In the large majority of cases, the pathological basis for the neck pain is unclear and the complaints are labeled as 'non-specific' or 'mechanical' [7]. Disability, limitations in activities and restrictions in participation in daily living and work, may be the result [30, 32]. The majority of the total costs of neck pain in the Netherlands were costs due to sick leave and disability payment [8].

Self-reported disability in patients with neck pain is often measured by means of region-specific questionnaires [25]. These questionnaires may measure disability or the functional status with greater responsiveness than generic health questionnaires [25]. Questionnaires should have good psychometric qualities, among which is reproducibility [25]. Reproducibility is the extent to which the same results are obtained on repeated tests when no real change in health status has occurred [14, 25]. Reproducibility may be influenced by random measurement errors and within patient variance [14, 27]. Both sources of variance may lead to score instability (natural variation) on repeated tests [14, 27]. If a patient with neck pain fills out the same questionnaire on two occasions, it is relevant to know what score instability can be expected in a predefined retest interval or in a waiting period. Reproducibility does have two aspects: reliability and agreement, representing test– retest score stability over time on group level and on individual level, respectively [14]. A measure for reliability is the intra-class correlation coefficient (ICC) [27]. It assesses not only the strength of the correlation between two repeated measures, but also if all measures on each subject are identical and do not differ systematically [27]. To quantify the agreement, the test–retest score stability over time on individual level, the limits of agreement (LOA) can be calculated according to the method of Bland and Altman [3, 6]. The LOA lies two standard deviations ($SD_{difference}$) above and under the mean total score difference of all patients between the first and second test. This means that, due to score instability, approximately 95% of all differences within patients will lie between these LOA on repeated tests [3, 6]. In an individual patient, the change due to treatment should exceed these LOA before one can state that real change has occurred.

The most used Neck Disability Questionnaires are the Neck Pain and Disability Scale (NPAD) [32] and the neck disability index (NDI) [30], which are both translated into several languages [1, 2, 5, 9, 13, 18–21, 23, 24, 28, 29, 33]. To be able to use the NPAD and NDI in different countries and social environments these questionnaires must not only be translated properly, but also culturally adapted and validated [4, 25]. These translations allow comparison of results of clinical research trials between countries. To investigate which questionnaire is most appropriate, psychometric studies are needed where questionnaires are applied simultaneously to the same sample of patients [25]. The advantage of the NPAD over the NDI may be the simple wording and the unitary structure of the questions; moreover, the questionnaire is easy to complete with its visual analog scale structure [9, 23, 25, 28]. The NPAD has not been formally translated into Dutch and, consequently, psychometric qualities of the Dutch Language Version (NPAD–DLV)

are unknown. The reliability of the NDI–DLV has been studied in patients with acute neck pain, but not in patients with chronic neck pain (CNP) other than patients with Whiplash associated disorder (WAD) [17, 31]. The first aim of this study was to translate the NPAD from English into Dutch. The second aim was to analyze test–retest reliability and agreement of the NPAD–DLV and the NDI–DLV in patients with CNP in an outpatient rehabilitation setting.

METHODS

Translation and cross-cultural adaptation of the NPAD

The NDAP was translated using a forward and backward translation procedure [4]. Two native Dutch speakers (a clinician, aware of the concepts behind the questionnaire and a staff member of the pain rehabilitation team) independently translated and culturally adapted the original version. The translated versions were critically reviewed reciprocally, compared with one another and with the original English version. Disagreements were discussed and a consensus version was produced. A backward translation (from Dutch into English) of this consensus version was made by a bilingual physiotherapist involved in spine research. The translators examined translation, backward translation and notes about the discussions made during the translation process. A concept version of the NPAD–DLV was developed and pilot-tested on a heterogeneous group of ten patients and employees of the rehabilitation center, who were asked to comment critically on understandability of the questions and instructions, responses, wording and layout. Finally, a final NPAD–DLV was produced.

Study sample

CNP patients were recruited from referrals by general practitioners or medical specialists for rehabilitation treatment in the Center for Rehabilitation at the University Medical Center Groningen, The Netherlands. Inclusion criteria for this study were: non-specific chronic neck pain ([3 months duration), admitted for rehabilitation, age between 18 and 65 years, less than 2 years out of work due to CNP or still at work with frequent sick leave due to neck pain, and sufficient knowledge of the Dutch language to complete questionnaires. Exclusion criteria were: specific neck pain, status post surgery in the cervical region, cardiovascular or pulmonary diseases significantly diminishing physical capacity, pregnancy, addiction to drugs, extensive psychological or behavioral problems.

Procedures

Prior to the first visit at the University Medical Center patients filled out a baseline questionnaire assessing demographics and clinical characteristics. During the first visit a review of the medical

history and a physical examination was performed. Immediately afterwards patients filled out the NPAD–DLV and the NDI–DLV. A second visit was scheduled, depending on subject availability, 1–5 weeks after the first visit, but prior to the start of the outpatient rehabilitation program. During the second visit the patients filled out the NPAD–DLV and the NDI–DLV for the second time. All patients signed informed consent prior to entering the study.

Measurements

The NPAD is a questionnaire whose development used the Million Visual Analogue Scale (VAS) as a template [16]. The NPAD consists of 20 items. Each item has a VAS of 100 mm with numeric anchors at 0, 1, 2, 3, 4 and 5 (each 20 mm apart). Item scores range from 0 (no pain or limitation in activities) to 5 (as much pain as possible or maximal limitation). The total NPAD score can vary from 0 to 100 points [16]. Test–retest reliability, expressed as intraclass correlation coefficient (ICC), of different language versions of the NPAD ranged from 0.81 to 0.98 with retest intervals from 1 day to 1–2 weeks [2, 9, 19, 21, 33].

The NDI is a questionnaire based on the Oswestry low back pain disability questionnaire and consists of 10 items [30]. Each item has six different assertions expressing progressive levels of pain or limitation in activities with a score between 0 (no pain or limitation) and 5 (as much pain as possible or maximal limitation). The total NDI scores can vary from 0 to 50 points [30]. Test–retest reliability, expressed as ICC, of different language versions of the NDI ranged from 0.50 to 0.97 with retest intervals from 1 day to 3 weeks [11, 12, 19, 20, 22, 24, 29, 31, 33]. The ICC of the NDI–DLV was 0.90 in patients with acute neck pain in general practice over a retest interval of 7 days [31]. For patients with WAD the test–retest reliability of the NDI–DLV was r = 0.81 over a retest interval of 3 months [17].

Data analyses

Descriptive statistics were calculated for the total scores of the two test sessions for both questionnaires. Reliability of the NDAP–DLV and the NDI–DLV was expressed as ICCs for the total scores. ICCs of 0.75 or higher were interpreted as acceptable reliability [27]. To quantify agreement (the test–retest score stability on individual level) of the NPAD–DLV and the NDI–DLV the limits of agreement were calculated as described by Bland and Altman [3, 6]. Statistical analyses were performed with SPSS 14.0.

RESULTS

A total of 181 neck patients were referred to the Center for Rehabilitation between November 2006 and December 2007. From this group 72 (40%) were admitted for rehabilitation. A total of 39 patients were eligible for inclusion in this study. During the waiting period after the first visit, 5

patients decided not to start with the rehabilitation program, 33 completed the NPAD–DLV and 32 the NDI–DLV twice. Characteristics of the study sample are presented in Table 2.1. In the translation and cross-cultural adaptation of the NPAD, minor changes were made in item 13 and item 18. In item 13 (outlook in life and the future), the given examples 'depression and hopelessness' were deleted because the respondents of the pre-final version found this superfluous. In item 18 (trouble with looking up or down) the text was changed into: bending the head forwards or backwards,

Table 2.1 Demographic and clinical characteristics of the study sample (n = 34)

	Mean (SD) or Median (IQR)	Min–Max
Age (years)	37.5 (11.5)	21–57
Duration of chronic pain (months)	12.5 (6.75–26.25)[a]	4–240
Sick leave in the past year (weeks)	21 (18.5)	0–52
	N	%
Female	23	68
Pain radiating to		
Shoulder(s)	30	88
Upper arm(s)	13	38
Forearm(s)	8	24
Hand/fingers	8	24
Between shoulder blades	15	44
Pins and needles below elbow	11	32
Concomitant complaints		
Headache	26	77
Dizziness	13	38
Concentration problems	8	24
Nausea	5	15
Fatigue	22	65
Low back pain	13	38
Self reported cause of neck pain		
Motor vehicle accident	19	56
Other trauma	1	12
Spontaneously/unknown	1	3
Stress	1	3
Work-related	2	6
Other	7	21
Previous treatment for neck pain	31	85
Education		
Low	2	6
Intermediate vocational education	23	68
High	7	21
Work status (self-employed/employee)	1/33	3/97
Sick leave	23	67
Involved in litigation	17	50

[a] Median and interquartile range for duration of pain (months).

Table 2.2 Total scores of the NPAD–DLV (n = 33) and NDI–DLV (n = 32) at test and retest, intraclass correlation coefficients (ICC), difference between test and retest and limits of agreement

	Test		Retest			Difference		
	Mean	SD	Mean	SD	ICC (95% CI)	Mean	SD	Limits of agreement
NPAD–DLV	50.7	15.7	51.2	14.7	0.76 (0.57 to 0.87)	1.3	10.4	± 20.9
NDI–DLV	22.6	5.9	20.6	6.4	0.84 (0.69 to 0.92)	-1.4	3.2	± 6.5

NPAD–DLV, Neck Pain and Disability Scale Dutch Language Version; NDI-DLV, Neck Disability Index Dutch Language Version.

because looking up and down can be done with the eyes only and without flexion–extension of the cervical spine. Details were added in the general instruction to emphasize that all items should be answered regarding the intensity of the neck pain or neck pain related disability.

Reliability and agreement

The mean retest interval was 18.2 days (SD 6.2, range 6–34). Test and retest results are presented in Table 2.2. Item 20 of the NPAD–DLV, concerning pain pills, was left blank by 2 (6%) patients. This item presumes the patient is taking medication. Item 7 of the NDI–DLV concerning driving was left blank by 2 (6%) patients; for these patients the score was adjusted using the mean of the answers on the rest of the questionnaire. The ICC of the NPAD–DLV was 0.76 (95% confidence interval (CI) 0.57–0.87) and of the NDI–DLV 0.84 (95% CI 0.69– 0.92). The LOAs of the NPAD–DLV and the NDI–DLV were, respectively, ±20.9 (scale 0–100) and ±6.5 (scale 0–50) (Table 2.2). Bland and Altman plots for the NPAD–DLV and the NDI–DLV are presented in Figures 2.1 and 2.2. No visible tendency towards unequal variance of the data appeared present.

DISCUSSION

The cross-cultural adaptation of the NPAD followed the forward and backward translation procedure. This procedure warranted on the one hand the meaning of the original items and on the other hand capturing of contents and meanings of the questions in the translation into the Dutch language. Other than the production of a more detailed general instruction, only minor modifications were made in items 13 and 18. The adaptation of item 18 is supported by the fact that the developers of the NPAD questionnaire have explicitly related this item to 'neck problems' [32] and other authors to 'neck dysfunction related to activities of the cervical spine' [2, 9, 13, 23]. The new NPAD–DLV was easy to comprehend. To complete both the NPAD– DLV and the NDI–DLV required maximally 15 min. The number of missing responses was negligible,

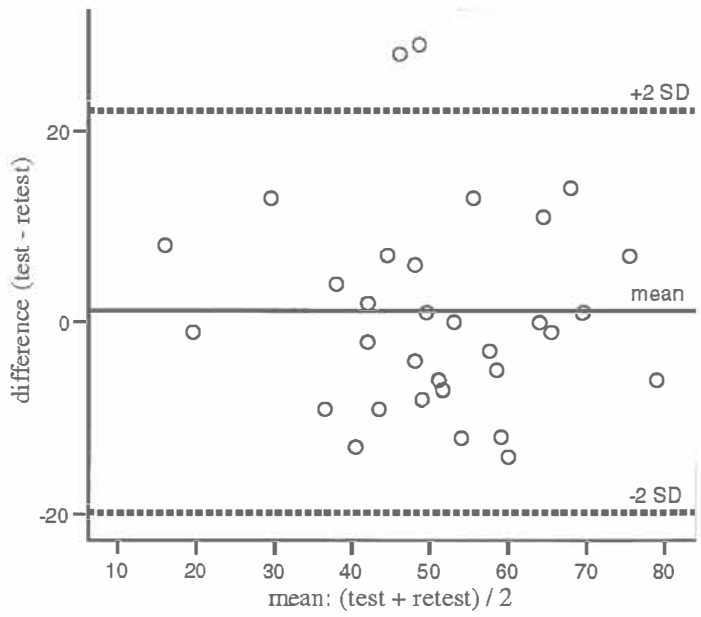


Figure 2.1 Mean score of each patient plotted against the difference between test-retest scores of NPAD–DLV. The reference line is the mean of total score difference of all patients. Limits of agreement at 2 SD.

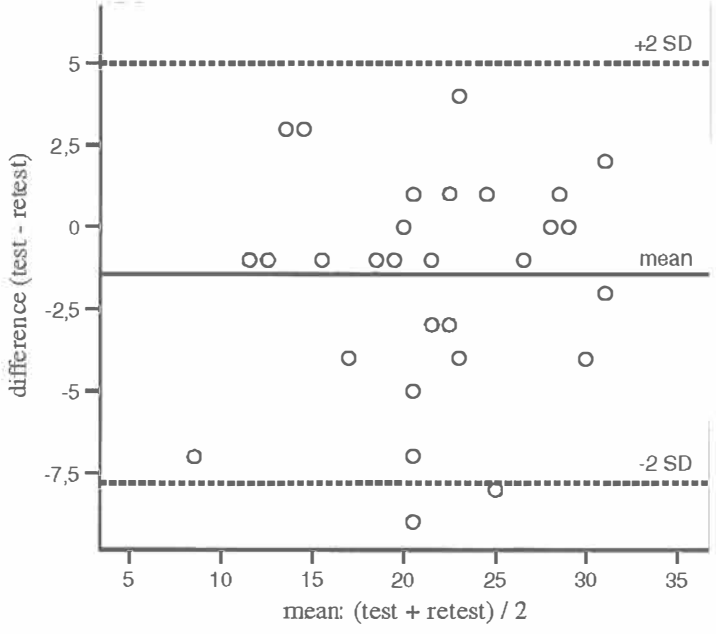


Figure 2.2 Mean score of each patient plotted against the difference between test-retest scores of NDI–DLV. The reference line is the mean of total score difference of all patients. Limits of agreement at 2 SD.

which was in agreement with other non-English versions of the questionnaires [1, 2, 5, 9, 13, 20, 21, 23, 24, 28, 29, 33].

The NPAD–DLV and the NDI–DLV demonstrated acceptable reliability in a sample of patients with CNP. The retest interval depended on the availability of the patient. In our rehabilitation setting it is normal to have a waiting period (1–2 months) between intake and start of the program. Therefore, it is interesting to know the extent of changes in questionnaire outcome occurring in absence of treatment. The sample sizes in our study were similar to other reproducibility studies [2, 9, 12, 19–22, 24, 29, 31, 33] except for four other studies [9, 12, 19, 33] where the sample sizes were 23, 17, 102 and 101, respectively. The female to male ratio in the current study is similar to that in most former reproducibility studies [9, 12, 19, 20, 22, 31, 33].

In our study, the mean total score of the NPAD–DLV was 50.7 and of the NDI–DLV 22.6. In other studies, the mean total scores of the NPAD ranged from 38.2 to 60.5 and of the NDI from 11.0 to 23.0 [2, 9, 11, 12, 19, 20, 22, 29, 31, 33]. In general, studies carried out in tertiary referral centers have higher total scores than those in primary care settings. In all studies, where both questionnaires were used, including ours, the NPAD scores were approximately 10% higher than the NDI scores when presented in % of a 0–100 scale [19, 21, 24, 33].

The reliability of the NPAD–DLV in our study (ICC = 0.76) was lower than in reliability studies with shorter retest intervals (less than 2 weeks ICC = 0.81–0.98 [2, 9, 19, 21, 33]). The reliability of the NDI–DLV in our study (ICC = 0.84) was somewhat lower than in most former NDI studies with generally shorter retest intervals (less than 2 weeks ICC = 0.50–0.97 [11, 19, 22, 24, 29, 31, 33]; 2 weeks, ICC = 0.88 [20]; 3 weeks ICC = 0.68 [12]). Perceived recovery (change) in the retest interval, to include the 'stable' patients in the reliability studies, was assessed in only the half of above mentioned NPAD and NDI studies [2, 29, 31, 33]. Apart from that it seems not to have resulted in differences in the extent of ICCs [2, 9, 20–22, 24, 29, 31, 33]. A trend may be seen that studies with a shorter retest interval do have higher ICCs. When looking at another region-specific questionnaire, the same trend was reported [10]. To test for the bias caused by differences in retest interval duration we assessed a partialled retest correlation; this means that we assessed the test–retest correlation for the NPAD–DLV and NDI–DLV while 'controlling' the effect of retest interval duration. The Pearson correlations for the test–retest reliability while 'controlling' or 'not controlling' for retest interval duration were r = 0.70 and r = 0.72 for the NPAD and r = 0.87 and r = 0.87 for the NDI. These results indicate that the influence of the effect of retest interval duration is minimal or negligible.

The NPAD is claimed to be a questionnaire with four underlying dimensions [32]. Factor analyses in other language NPADs identified two to four factors on which different items were loading [9, 13, 23, 24, 33]. The factorial structure presented in the original publication was based on a relatively small sample (n = 95); therefore, the stability of the observed factor solution may be questioned and too sample specific to be reproducible in different samples. Therefore, comparison of the ICCs for subscales in different (language) NPAD studies is challenging.

However, a principal component analysis in a German study with a sample size of 448 indicated a one-factor solution for the NPAD, and it was concluded that the NPAD is a multidimensional assessment instrument measuring different dimensions of one construct neck pain, in a stable manner [28]. Because above mentioned reasons and because factor analysis was not an aim of the current study only total scores were used to analyze the reliability of the NPAD–DLV.

If a patient with neck pain fills out the same questionnaire on two occasions, in a waiting period prior to the start of a rehabilitation program, a (very) short time interval increases the probability of carryover or recall effects due to memory, mood or practice, whereas a larger interval increases the probability the clinical status has changed and that the score of the first session has been forgotten [27]. There are several explanations for possible changes of the clinical status during the waiting period: the effect of the clinic consultation, the anticipation of the patient on the program, the effect of a period of waiting before the real rehabilitation program starts, the chronic neck pain itself with its fluctuations and the questionnaires itself [22].

To quantify the agreement, the test–retest score stability over time on individual level, the 'limits of agreement' (LOA) were calculated. No criteria are available for interpretation of the LOA. Smaller LOA means more stability and indicate that the natural variation is smaller. The $SD_{difference}$ (the standard deviation of the mean total score difference off all patients between the first and second test) and the LOA of the NPAD–DLV in the present study (10.4 and ±20.9) were somewhat higher than in one other study (9.0 and ±17.9) with a retest interval of 1 day [33]. In this French study, a 5 point ordinal transition scale was used to include clinically stable patients. Despite a clear difference in retest intervals, the differences in $SD_{difference}$ were small. The $SD_{difference}$ and the LOA of the NDI–DLV in the present study (3.2 and ±6.5) was similar to most other NDI studies with shorter retest intervals (1 day, $SD_{difference}$ 3.4, LOA ±6.7 [33]; 1 week, $SD_{difference}$ 3.9, LOA ±7.8 [31]; 1 week $SD_{difference}$ 1.5, LOA ±3.0 [29]; 1–2 weeks, $SD_{difference}$ 4.4, LOA ±8.9 [22]). In three of these four studies, a transition scale was used to include clinically stable patients [22, 29, 31, 33]. Proportionally, the $SD_{difference}$ of the Greek study [29] was similar to the $SD_{difference}$ of the present study ($SD_{difference}$/mean total score was 0.12 and 0.14, respectively). The NDI reliability studies have shown smaller $SD_{difference}$ and smaller natural variations compared to the NPAD reliability studies [22, 29, 31, 33]. Larger instability of the NPAD may be explained by differences in operationalizations of 'neck disability' between items of the NPAD and the NDI [30, 32]. Post hoc analysis showed that the amount of natural variation of the NPAD–DLV could not be attributed to individual items of the questionnaire. Clinical effects of therapy in an individual patient should exceed the limits of agreement before one can state that real change has occurred. The minimal clinically important difference (MCID) is suggested to be 11 points on the NPAD [9] and 2 to 10 points on the NDI [11, 12, 26, 31]. Based on the variation in the current study, patients have to change at least 21 points on the NPAD–DLV (scale 0–100) and at least 7 points on the NDI–DLV (scale 0–50), will these patients be judged as having 'really' changed.

There are limitations to consider in evaluating our research. First, the sample size is relatively small (n<50), therefore, our sample could have misestimated the "true" population ICCs and LOAs. However, the ICCs of the NPAD–DLV and NDI–DLV were in line with two studies with larger samples and with ICCs of, respectively, 0.81 and 0.86 (n = 102) [19] and 0.91 and 0.93 (n = 101) [33]. Second, the retest interval was longer than in most other NPAD and NDI studies [2, 9, 11, 19–22, 24, 29, 31, 33]. Therefore, the reported ICC and LOA values in the present study probably underestimate the reliability and agreement of the scales. Nevertheless the retest interval duration may be not the only important factor influencing these values, in view of an NDI study with a retest interval of 2.5 days (SD ± 0.95) where the ICC was 0.50 [11]. Other factors such as symptom duration (acute, sub-acute or chronic), patient setting (primary, secondary or tertiary care) and mean disability score on the questionnaires may also influence the values of ICC and LOA. Third, the retest interval was not fixed and perceived recovery (change) in this interval was not controlled for. However, $SD_{difference}$ of the NDI–DLV in our sample was similar to studies where change was controlled for.

A strength of this study is that to the authors' knowledge for the second time a reproducibility study is made for the NPAD with respect to reliability and limits of agreement with a head to head comparison with the NDI. Further study with the NPAD–DLV is necessary to assess the reliability and agreement in other patient groups (e.g. acute, sub-acute, primary care patients), to assess the ICC with a shorter retest interval and to study other measurement properties, such as validity, responsiveness and MCID.

Conclusion

A reliable DLV of the NPAD was developed. The reliability of the NPAD–DLV and the NDI–DLV was acceptable for patients with CNP within an outpatient rehabilitation setting. The natural variation ('instability') in the NPAD–DLV total scores was relatively large and larger than the variation of the NDI–DLV.

Acknowledgements

The authors thank Marleen Speller for assistance in data collection.

REFERENCES

1. Ackelman BH, Lindgren U (2002) Validity and reliability of a modified version of the neck disability index. J Rehabil Med 34:284–287

2. Agarwal S, Allison GT, Agarwal A, Singer KP (2006) Reliability and validity of the Hindi version of the Neck Pain and Disability Scale in cervical radiculopathy patients. Disabil Rehabil 28:1405–1411

3. Altman DG, Bland JM (1983) Measurement in Medicine: The analysis of method comparison studies. Statistician 32:307–317

4. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25:3186–3191

5. Bicer A, Yazici A, Camdeviren H, Erdogan C (2004) Assessment of pain and disability in patients with chronic neck pain: reliability and construct validity of the Turkish version of the neck pain and disability scale. Disabil Rehabil 26:959–962

6. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310

7. Bogduk N (1999) The neck. Baillieres Clin Rheumatol 13:261–285

8. Borghouts JA, Koes BW, Vondeling H, Bouter LM (1999) Cost-of-illness of neck pain in The Netherlands in 1996. Pain 80:629–636

9. Bremerich FH, Grob D, Dvorak J, Mannion AF (2008) The Neck Pain and Disability Scale: cross-cultural adaptation into German and evaluation of its psychometric properties in chronic neck pain and C1-2 fusion patients. Spine 33:1018–1027

10. Brouwer S, Kuijer W, Dijkstra PU, Goeken LN, Groothoff JW, Geertzen JH (2004) Reliability and stability of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. Disabil Rehabil 26:162–165

11. Cleland JA, Childs JD, Whitman JM (2008) Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. Arch Phys Med Rehabil 89:69–74

12. Cleland JA, Fritz JM, Whitman JM, Palmer JA (2006) The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine (Phila Pa 1976) 31:598–602

13. Cook C, Richardson JK, Braga L, Menezes A, Soler X, Kume P, Zaninelli M, Socolows F, Pietrobon R (2006) Cross-cultural adaptation and validation of the Brazilian Portuguese version of the Neck Disability Index and Neck Pain and Disability Scale. Spine 31:1621–1627

14. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ (2001) Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 17:479–487

15. Ferrari R, Russell AS (2003) Regional musculoskeletal conditions: neck pain. Best Pract Res Clin Rheumatol 17:57–70

16. Goolkasian P, Wheeler AH, Gretz SS (2002) The neck pain and disability scale: test-retest reliability and construct validity. Clin J Pain 18:245–250

17. Heymans WFGJ, Lutke Schipholt HJA, Elvers JWH (2002) Neck Disability Index Dutch Version (NDI-DV): investigation of reliability in patients with chronic whiplash. Ned T Fysiother 112:94–99

18. Köke AJA, Heuts PHTG, Vlaeyen JWS (1996) Neck Disability Index. Pain Knowledge Center, University Medical Center Maastricht, Measurement instruments chronic pain. (Pijn Kennis Centrum, Academisch Ziekenhuis Maastricht, Meetinstrumenten chronische pijn), Maastricht

19. Kose G, Hepguler S, Atamaz F, Oder G (2007) A comparison of four disability scales for Turkish patients with neck pain. J Rehabil Med 39:358–362

20. Kovacs FM, Bago J, Royuela A, Seco J, Gimenez S, Muriel A, Abraira V, Martin JL, Pena JL, Gestoso M, Mufraggi N, Nunez M, Corcoll J, Gomez-Ochoa I, Ramirez MJ, Calvo E, Castillo MD, Marti D, Fuster S, Fernandez C, Gimeno N, Carballo A, Milan A, Vazquez D, Canellas M, Blanco R, Brieva P, Rueda MT, Alvarez L, Del Real MT, Ayerbe J, Gonzalez L, Ginel L, Ortega M, Bernal M, Bolado G, Vidal A, Ausin A, Ramon D, Mir MA, Tomas M, Zamora J, Cano A (2008) Psychometric characteristics of the Spanish version of instruments to measure neck pain disability. BMC Musculoskelet Disord 9:42

21. Lee H, Nicholson LL, Adams RD, Maher CG, Halaki M, Bae SS (2006) Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. Spine 31:1841–1845

22. McCarthy MJ, Grevitt MP, Silcocks P, Hobbs G (2007) The reliability of the Vernon and Mior neck disability index, and its validity compared with the short form-36 health survey questionnaire. Eur Spine J 16:2111–2117

23. Monticone M, Baiardi P, Nido N, Righini C, Tomba A, Giovanazzi E (2008) Development of the Italian version of the Neck Pain and Disability Scale, NPDS-I: cross-cultural adaptation, reliability, and validity. Spine 33:E429–E434

24. Mousavi SJ, Parnianpour M, Montazeri A, Mehdian H, Karimi A, Abedi M, Ashtiani AA, Mobini B, Hadian MR (2007) Translation and validation study of the Iranian versions of the Neck Disability Index and the Neck Pain and Disability Scale. Spine 32:E825–E831

25. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine 27:515–522

26. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC (2007) Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine (Phila Pa 1976) 32:3047–3051

27. Portney LG, Watkins MP (2000) Foundations of Clinical Research 2nd Edition Prentice Hall Health. , Upper Saddle River, New Yersey 07458

28. Scherer M, Blozik E, Himmel W, Laptinskaya D, Kochen MM, Herrmann-Lingen C (2008) Psychometric properties of a German version of the neck pain and disability scale. Eur Spine J 17:922–929

29. Trouli MN, Vernon HT, Kakavelakis KN, Antonopoulou MD, Paganas AN, Lionis CD (2008) Translation of the Neck Disability Index and validation of the Greek version in a sample of neck pain patients. BMC Musculoskelet Disord 9:106

30. Vernon H, Mior S (1991) The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 14:409–415

31. Vos CJ, Verhagen AP, Koes BW (2006) Reliability and responsiveness of the Dutch version of the Neck Disability Index in patients with acute neck pain in general practice. Eur Spine J 15:1729–1736

32. Wheeler AH, Goolkasian P, Baird AC, Darden BV (1999) Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. Spine 24:1290–1294

33. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel M (2002) French translation and validation of 3 functional disability scales for neck pain. Arch Phys Med Rehabil 83:376–382

Chapter 3

Neck Pain and Disability Scale and Neck Disability Index: validity of Dutch language versions

Wim Jorritsma
Grietje E. de Vries
Pieter U. Dijkstra
Jan H. B. Geertzen
Michiel F. Reneman

Eur Spine J (2012) 21:93–100

ABSTRACT

Purpose: To investigate the validity of the Neck Pain and Disability Scale Dutch Language Version (NPAD–DLV) and the Neck Disability Index (NDI)–DLV.

Methods: NPAD–DLV, NDI–DLV, Short-Form-36 Health Survey (SF-36)–DLV, visual analog scale $(VAS)_{pain}$ and $VAS_{disability}$ were administered to 112 patients with non-specific chronic neck pain in an outpatient tertiary rehabilitation setting. Twenty seven hypotheses were formulated regarding validity. NPAD–DLV and NDI–DLV were evaluated for content validity (normal distribution total scores, missing items, floor and ceiling effects), internal consistency (Cronbach's alpha and Spearman Item-total correlations), construct validity (Pearson correlations with SF-36 domains, VAS_{pain} and $VAS_{disability}$ and Pearson correlation between total scores of NPAD–DLV and NDI–DLV).

Results: NPAD–DLV and NDI–DLV scores were distributed normally. Missing items were negligible. Floor and ceiling effects were absent in NPAD–DLV and in NDI–DLV two items had floor effects and one item had a ceiling effect. Cronbach's alpha of NPAD–DLV was 0.93 and of NDI–DLV 0.83. Item-total correlations ranged for NPAD-DLV from 0.45 to 0.73 and for NDI–DLV from 0.40 to 0.64. The correlation between, respectively, NPAD–DLV and NDI–DLV and: SF-36 domains ranged from -0.36 to -0.70 and from -0.34 to -0.63; VAS_{pain} was 0.54 and 0.43; $VAS_{disability}$ was 0.57 and 0.52. The correlation between the total scores of NPAD–DLV and NDI–DLV was 0.77. Twenty six hypotheses were not rejected and one hypothesis was rejected.

Conclusion: The NPAD–DLV and NDI–DLV are valid measures of self-reported neck-pain related disability.

INTRODUCTION

Neck pain is a common musculoskeletal complaint in western societies [11]. In the majority of cases the pathological basis for neck pain is unclear and complaints are labeled as 'non-specific' or 'mechanical' [4]. Neck pain may result in disability, limitations in activities and restrictions in participation in daily living and work [32, 34]. Self-reported disability in patients with neck pain is often measured by means of region-specific and generic questionnaires [25]. Questionnaires should have good psychometric qualities, including validity [25, 27]. Three aspects of validity will be tested in this study. Content validity is the extent to which items of the questionnaire reflect all aspects of the construct to be measured [25, 27]. Internal consistency is the extent to which all items measure the same construct [25, 27]. Construct validity is the extent to which a questionnaire is convergent and/or divergent correlated with other tests that are presumed to measure a similar or different construct [25, 27].

The most frequently used neck disability questionnaires are the Neck Pain and Disability Scale (NPAD) [34] and the Neck Disability Index (NDI) [32], which are validated in several languages [2, 3, 6, 8, 16, 17, 19, 21, 22, 29, 35]. The validity of the Dutch Language Versions (DLV) of the NPAD and NDI has not been studied. The aim of this study was to investigate the validity of the NPAD–DLV and the NDI–DLV in patients with non-specific chronic neck pain (CNP) in an outpatient tertiary rehabilitation setting. A priori hypotheses were defined (Text box 3.1) and outlined in "Materials and methods".

MATERIALS AND METHODS

Study sample

Patients with CNP were recruited from referrals by general practitioners or medical specialists for rehabilitation treatment in the Center for Rehabilitation at the University Medical Center Groningen, The Netherlands. Inclusion criteria for this study were non-specific chronic neck pain (>3 months duration), admitted for outpatient rehabilitation, age between 18 and 65 years, and sufficient knowledge of the Dutch language (to complete questionnaires). Exclusion criteria were status post surgery in the cervical region, cardiovascular or pulmonary diseases severely diminishing physical capacity, pregnancy, addiction to drugs, and extensive psychological or behavioral problems.

Procedures

Prior to the first visit patients filled out a baseline questionnaire assessing clinical characteristics including visual analog scale $(VAS)_{pain}$ and $VAS_{disability}$. During the first visit a review of the medical history and a physical examination was performed. A second visit was scheduled, depending

on the length of the waiting list and patient availability, 2–9 weeks after the first visit, but prior to the start of the rehabilitation program. During the second visit the patients filled out the NPAD–DLV, the NDI–DLV and the Short-Form-36 Health Survey (SF-36). All patients signed informed consent for their data to be used for research purposes. Data were gathered between November 2006 and October 2009.

Measurements

The NPAD consists of 20 items divided into 4 dimensions; neck problems; pain intensity; emotion and cognition; and interference with life activities [34]. Each item has a VAS of 100 mm with numeric anchors at 0, 1, 2, 3, 4 and 5 (each 20 mm apart). Item scores range from 0 (no pain or activity limitation) to 5 (as much pain as possible or maximal limitation). The total NPAD score ranges from 0 to 100 points. Higher scores indicate greater disability [34]. The NPAD has shown to be a valid and responsive measure of disability in other languages [3, 6, 8, 17, 19, 21, 22, 29, 34, 35]. The NPAD–DLV was used in this study; the reproducibility is acceptable [15].

The NDI consists of ten items: pain intensity, personal care, lifting, reading, headaches, concentration, work, driving, sleeping, and recreation [32]. Each item has six different assertions expressing progressive levels of pain or limitation in activities. Item scores range from 0 (no pain or limitation) to 5 (as much pain as possible or maximal limitation). The total NDI score ranges from 0 to 50 points. Higher scores indicate greater disability [32]. The NDI has shown to be a valid and responsive measure of disability in different languages [2, 8, 17, 19, 20, 22, 26, 32, 33, 35]. The NDI–DLV [16] was used in this study; the reproducibility [15, 33] and responsiveness are acceptable [26, 33].

The SF-36 is a questionnaire assessing general health of the past 4 weeks in 8 domains: physical functioning, physical role restriction, bodily pain, general health, vitality, social functioning, emotional role restriction, and mental health [12]. Scores for each domain range from 0 to 100, with higher scores indicating higher levels of functioning or well-being. The Dutch language version of the SF-36 has shown to be reliable and valid [1].

The VAS_{pain} is a horizontal line, 100 mm in length, anchored by word descriptors at each end (0: no pain, 100: worst pain possible). Patients are asked to draw a vertical mark across the horizontal line that best represents the pain level. The VAS_{pain} is a commonly used assessment instrument with proven reliability and validity [9].

The $VAS_{disability}$ was evaluated by the question 'how much does your neck pain restrict you in your daily activities?' (ADL, housekeeping, work, hobby, recreation, sport and social activities). The scoring procedures are similar to the VAS_{pain}. The anchoring word descriptors are 0: no restriction and 100: worst possible restriction. The reliability and validity of the $VAS_{disability}$ were assessed in patients with chronic musculoskeletal pain [5].

Hypotheses

Hypotheses are listed in Text box 3.1 and for the most part based on previous studies as described below.

Content validity

A normal distribution of the total scores of the NPAD–DLV and NDI–DLV was expected (Hypothesis 1), a good completeness of item responses (Hypothesis 2), and no floor and ceiling effects in item responses were expected (Hypothesis 3) [6, 7, 17, 19, 34, 35]. It was expected that scores on the NDI in a tertiary rehabilitation setting would be significantly higher than those in a Dutch primary care setting (Hypothesis 4) [6, 14, 19, 20, 26, 33, 34]. No Dutch data are available for comparison of the NPAD–DLV.

Text box 3.1 Hypothesis for examining validity of the NPAD–DLV and NDI–DLV

The validity is not rejected when:

Content validity

1. The total scores are normally distributed
2. The percentage of missing items is <5%
3. Floor and ceiling effects in item responses are not present
4. Total scores on the NDI of patients with CNP in a tertiary rehabilitation setting are significantly higher than patients with neck pain in a primary care setting

Internal consistency

5. The Cronbach's alphas are ≥0.7
6. The strength of the relationship of the single items with the total scale is fair to moderate ($0.25 \leq r < 0.75$)

Construct validity

7. The strength of the relationship with all eight SF-36 domains is fair to moderate ($0.25 \leq r < 0.75$)
8. The strength of the relationship with VAS_{pain} is fair to moderate ($0.25 \leq r < 0.75$)
9. The strength of the relationship between the NPAD and VAS_{pain} is higher than the strength of the relationship between the NDI and VAS_{pain}
10. The strength of the relationship with $VAS_{disability}$ is moderate ($0.50 \leq r < 0.75$)
11. Differences in total scores between two age groups (below and above mean age of the study population) are not significant
12. Differences in total scores between males and females are not significant
13. Total scores of patients who are in litigation because of CNP are significantly higher than patients who are not in litigation
14. Total scores of patients who are receiving workers compensation because of CNP are significantly higher than patients who are not receiving workers compensation

All hypotheses are operative for both questionnaires with exception of hypotheses 4, 9 and 15.

Internal consistency

It was expected that Cronbach's alphas of the NPAD–DLV and NDI–DLV would be ≥0.70 (Hypothesis 5) and that Item–total score correlations would be fair to moderate (Hypothesis 6) [6, 8, 17, 19–22, 29, 32, 34, 35].

Construct validity

A fair to moderate correlation with all eight SF-36 domains was expected (Hypothesis 7) [8, 20–22]. It was expected that the NPAD–DLV and NDI–DLV had a fair to moderate correlation with VAS_{pain} [2, 3, 7, 13, 17, 20–22, 35] and a moderate correlation with $VAS_{disability}$ [17, 35] (Hypotheses 8 and 10). Because four questions of the NPAD are pain-oriented a stronger correlation between the NPAD–DLV and the VAS_{pain} was expected than between the NDI–DLV and the VAS_{pain} (Hypothesis 9). No significant differences between sexes or age groups (below and above mean age of the study population) were expected (Hypotheses 11 and 12) [20, 32]. Significantly higher NPAD–DLV and NDI–DLV scores were expected for patients who were in litigation or who were receiving workers compensation because of their neck problems than for patients who were not in litigation or who received no workers compensation (Hypotheses 13 and 14) [18, 28]. A moderate-to-good correlation between the total scores of the NPAD–DLV and NDI–DLV was expected (Hypothesis 15) [2, 10, 22, 35]. All hypotheses are operative for both the NPAD–DLV and NDI–DLV with exception of hypotheses 4, 9 and 15; in total this results in 27 hypotheses.

Data analyses and criteria

Normality of the total scores was analyzed using the Kolmogorov–Smirnov test and PP plots. Floor and ceiling effects were considered to be present if more than 15% of respondents achieved the lowest or highest possible score for items [6]. When ≥75% of the items did not have floor or ceiling effects, these questionnaires were considered to have no floor or ceiling effects. Internal consistency was assessed with Cronbach's alphas and values ≥0.7 are considered adequate [24]. Standardized item-total score Spearman correlations of the NPAD–DLV and NDI–DLV were analyzed by calculating correlation coefficients between each item and the sum of all other items excluding the item investigated. Independent t-tests were used to analyze differences NPAD–DLV and NDI–DLV total scores between tertiary and primary care patients, patients younger or older than the mean age of the study population, men and women, patients with or without litigation, and with or without workers compensation. Pearson correlations were used to determine the strength of the relationship between the total scores of the NPAD–DLV and NDI–DLV and the SF-36 domain scores, VAS_{pain} and $VAS_{disability}$ and also between the total scores of NPAD–DLV and NDI–DLV. The construct validity was interpreted as good when at least 75% of the results corresponded with the hypotheses [30]. Correlations were interpreted as follows: $0.75 \leq r \leq 1.0$ as good, $0.50 \leq r < 0.75$ moderate, $0.25 \leq r < 0.50$ fair, and $0.00 \leq r < 0.25$ little or no [27].

All statistical analyses were performed with SPSS software, version 16.0. The critical values for significance were set at $p<0.05$.

RESULTS

A total of 391 patients with CNP were referred to the Center for Rehabilitation between November 2006 and October 2009 of which 129 were admitted for rehabilitation. A total of 125 patients fulfilled inclusion criteria. During the waiting period 13 patients decided not to start with the rehabilitation program because of lack of time, waiting period too long, problems with insurance company, and further diagnostic procedures. Clinical characteristics of the patients (n = 112) are presented in Table 3.1.

Content validity

NPAD–DLV and NDI–DLV were normally distributed. Therefore, hypothesis 1 was not rejected. Mean scores for individual items for the NPAD–DLV ranged from 1.7 to 4.2 (Table 3.2) and for the NDI–DLV from 0.7 to 2.8 (Table 3.3). In total 22 (1%) of 2,240 NPAD–DLV items and 15 (1%) of 1,120 NDI–DLV items were missing; therefore hypothesis 2 was not rejected (Tables 3.2, 3.3). Floor effects were <10% for all NPAD–DLV items. Ceiling effects were <13% for all NPAD–DLV items; therefore hypothesis 3 was not rejected (Table 3.2). For the NDI–DLV the items 'personal care' and 'sleeping' had floor effects, with respectively 44 and 19% of the patients scoring the lowest possible value. A ceiling effect was present for 'headaches' (19% of patients scored highest). Because 8 out of 10 NDI–DLV items did not have floor effects and 9 out of 10 did not have ceiling effects, hypothesis 3 was not rejected (Table 3.3). The total NDI–DLV score was 21.5 (Table 3.1). This score is significantly higher than the total scores in a Dutch primary care setting ($t(293) = 8.2$ (95% CI 5.3–8.7) [26] and $t(297) = 8.3$, (95% CI 5.3–8.7) [33]); therefore hypothesis 4 was not rejected.

Internal consistency

The Cronbach's alphas of the NPAD–DLV and the NDI–DLV were respectively 0.93 and 0.83; therefore hypothesis 5 was not rejected. The strength of all Item–total correlations ranged from $r = 0.45$ to $r = 0.73$ (NPAD–DLV) and from $r = 0.40$ to $r = 0.64$ (NDI–DLV) (Tables 3.2, 3.3). Because all Item–total correlations fell within the hypothesized ranges, hypothesis 6 was not rejected.

Construct validity

Correlations between the total scores and SF-36, VAS_{pain}, and $VAS_{disability}$ are presented in Table 3.4. Differences between age groups, sexes, litigation status, and workers' compensation are

Table 3.1 Patient characteristics (n = 112)

	Mean (SD) n (%)
Age (years)	38.8 (11.4)
Duration of chronic pain (months)	18.0 (8.0–48.0)[a]
Sick leave in the past year (weeks)	15.6 (18.1)
NPAD-DLV (scale 0–100)	53.1 (16.6)
NDI-DLV (scale 0–50)	21.5 (7.4)
VAS_{pain} (0–100)	53.2 (21.4)
$VAS_{disability}$ (0–100)	54.0 (23.5)
Female	70 (63)
Pain radiating to	
Shoulder(s)	94 (84)
Upper arm(s)	55 (49)
Forearm(s)	36 (32)
Hand/fingers	33 (30)
Between shoulder blades	54 (50)
Pins and needles below elbow	36 (34)
Concomitant complaints	
Headache	81 (73)
Dizziness	38 (34)
Concentration problems	20 (18)
Nausea	13 (12)
Fatigue	69 (62)
Low back pain	44 (40)
Self reported cause of neck pain	
Motor vehicle accident	47 (42)
Other trauma	16 (14)
Spontaneously/unknown	11 (10)
Stress	5 (5)
Work-related	12 (11)
Other	21 (19)
Previously been treated for neck pain	102 (92)
Education	
Low	4 (4)
Intermediate vocational education	82 (75)
High	23 (21)
Work status (employed)	94 (84)
Workers compensation	62 (55)
Involved in litigation	34 (31)

NPAD–DLV, Neck Pain and Disability Scale Dutch Language Version; NDI–DLV, Neck Disability Index Dutch Language Version; VAS, Visual Analog Scale.
[a] Median and Interquartile Range.

presented in Table 3.5. Hypotheses 7–13 were not rejected. Hypothesis 14 was rejected for the NPAD–DLV and not rejected for the NDI–DLV. The relationship between total scores of NPAD–DLV and NDI–DLV is presented in Figure 3.1. The strength of the correlation between the NPAD–DLV and NDI–DLV was r = 0.77 (Table 3.4); therefore hypothesis 15 was not rejected.

DISCUSSION

In this study the validity of the DLV of the NPAD and the NDI was tested with the use of pre-defined hypotheses. Because 26 of the 27 (96%) pre-defined hypotheses were not rejected the validity of the NPAD–DLV and NDI–DLV was interpreted as good. The current study was conducted in a university setting and is therefore representative for patients with CNP in a tertiary

Table 3.2 Descriptive data and distribution of responses for each item in the NPAD–DLV (n = 112) and Spearman correlation between item scores and total score

	Item	Present study Mean (SD)	% Individuals with lowest score	% Individuals with highest score	Numbers of missing	Item–total correlation
1.	Pain intensity	2.6 (1.0)	1	2	1	.66
2.	Average pain	2.9 (0.9)	0	1	1	.58
3.	Worst pain	4.2 (0.7)	0	13	1	.45
4.	Sleeping	2.5 (1.5)	9	3	1	.60
5.	Standing	2.1 (1.2)	2	1	1	.66
6.	Walking	2.0 (1.1)	5	1	1	.68
7.	Driving/riding	2.6 (1.3)	5	2	1	.67
8.	Social activities	2.9 (1.2)	3	4	1	.73
9.	Recreational activities	3.0 (1.1)	2	2	1	.72
10.	Working	3.4 (1.1)	1	7	1	.57
11.	Personal care	1.7 (1.3)	10	0	1	.70
12.	Personal relationships	2.1 (1.3)	6	0	1	.70
13.	Outlook on life and future	2.2 (1.6)	9	5	1	.57
14.	Emotions	2.5 (1.4)	4	1	1	.61
15.	Thinking/concentration	2.7 (1.5)	7	4	1	.55
16.	Neck stiffness	2.5 (1.2)	3	1	1	.50
17.	Turning neck	2.4 (1.3)	5	1	1	.57
18.	Looking up/down	2.3 (1.4)	6	2	2	.52
19.	Working overhead	3.5 (1.2)	1	8	1	.50
20.	Effect of pain pills	2.7 (1.4)	5	6	2	.47

All correlations significant at the 0.01 level (2-tailed).

referral center. The sample size in our study was similar to those of other validity studies [6, 17, 35]. In the current study more women (63%) than men (37%) were included; this is similar to other validity studies [6, 10, 14, 17, 20–22, 29, 32, 34, 35], where the female proportions ranged from 54 to 83%. The mean age in our study was relatively young (39 years) in comparison with

Table 3.3 Descriptive data and distribution of responses for each item in the NDI–DLV (n = 112) and Spearman correlations between item score and total score

Item		Present study Mean (SD)	% Individuals with lowest score	% Individuals with highest score	Numbers of missing	Item–total correlation
1.	Pain	2.2 (0.8)	0	1	1	.48
2.	Personal Care	0.7 (0.8)	44	0	1	.50
3.	Lifting	2.3 (1.3)	5	2	1	.50
4.	Reading	2.4 (1.0)	6	0	1	.40
5.	Headaches	2.8 (1.6)	13	19	1	.54
6.	Concentration	2.0 (1.3)	12	2	2	.53
7.	Work	2.5 (1.2)	9	6	2	.55
8.	Driving	2.4 (1.2)	8	3	2	.48
9.	Sleeping	1.9 (1.3)	19	1	2	.56
10.	Recreation	2.4 (1.0)	3	1	2	.64

All correlations significant at the 0.01 level (2-tailed).

Table 3.4 Construct validity of the NPAD–DLV and NDI–DLV (Pearson correlations)

	NPAD	95% CI	NDI	95% CI
NDI	.77	.68 to .84	-	
VAS_{pain}	.54	.39 to .66	.43	.27 to .57
$VAS_{disability}$	.57	.43 to .68	.52	.37 to .64
SF-36 Physical functioning	-.58	-.69 to -.44	-.49	-.62 to -.33
SF-36 Role physical	-.36	-.51 to -.19	-.38	-.53 to -.21
SF-36 Bodily pain	-.70	-.78 to -.59	-.63	-.73 to -.50
SF-36 General health	-.44	-.58 to -.28	-.47	-.60 to -.31
SF-36 Vitality	-.50	-.63 to -.35	-.51	-.64 to -.36
SF-36 Social functioning	-.58	-.69 to -.44	-.61	-.71 to -.48
SF-36 Role emotional	-.39	-.54 to -.22	-.37	-.52 to -.20
SF-36 Mental health	-.45	-.59 to -.29	-.34	-.49 to -.16

NPAD–DLV, Neck Pain and Disability Scale Dutch Language Version; NDI–DLV, Neck Disability Index Dutch Language Version; VAS, Visual Analog Scale; SF-36, Short Form Health Survey.
All correlations significant at the 0.01 level (2-tailed).

Table 3.5 Results of independent t-tests for the comparison of age ≤39 versus age >39, male versus female, litigation versus no litigation, workers compensation (WC) versus no WC

		NPAD			NDI		
		Mean (SD)	95% CI	p-value	Mean (SD)	95% CI	p-value
Age	≤39	51.1 (15.8)	-10.63 to 1.86	.167	20.7 (6.3)	-4.62 to 1.11	.227
	>39	55.5 (17.3)			22.4 (8.5)		
Gender	Male	54.3 (17.5)	-4.57 to 8.36	.562	22.6 (8.2)	-1.12 to 4.65	.228
	Female	52.4 (16.1)			20.8 (6.9)		
Litigation[a]	Yes	57.6 (16.9)	-13.68 to -3.5	.020	25.4 (6.8)	-8.76 to -3.09	<.001
	No	50.5 (15.8)			19.5 (6.9)		
WC[a]	Yes	55.3 (15.8)	-11.17 to 1.35	.062	23.5 (6.9)	-7.21 to -1.82	<.001
	No	50.4 (17.2)			19.0 (7.4)		

NPAD–DLV, Neck Pain and Disability Scale Dutch Language Version; NDI–DLV, Neck Disability Index Dutch Language Version; CI, confidence interval.
[a] 1-tailed.

other validity studies, where the mean age of patients ranged from 38 to 65 years [6, 10, 14, 17, 20–22, 29, 32, 34, 35].

The normality of the total scores and the completeness of item responses were similar to other studies [6, 19, 21, 29, 35]. Floor and ceiling effects were not found in two studies [22, 35], while in two other studies floor effects for NPAD (6 items [19] and 14 items [6]) and NDI (3 items [19]) and ceiling effects for the NDI (1 item [19]) were found. The lower scores for most of the items for the NPAD [6, 19] and NDI [6] in those studies may explain the differences in floor effects with the present study. It is of interest that in the German study [6] (n = 108 of which n = 80 after atlantoaxial screw fixation and n = 28 with CNP) in the subgroup of patients with CNP much less items (3 in stead of 14) had floor effects. The Korean study [19] (n = 180) consisted of patients treated in physiotherapy departments of private hospitals or clinics.

We calculated a single Cronbach's alpha for the NPAD–DLV and NDI–DLV because their factor structure (1, 2, 3, or 4 factors for NPAD and 1 or 2 factors for NDI) is unclear and because in the original English versions single Cronbach's alphas for the total scales were calculated [6, 8, 21–23, 29, 31, 32, 35]. In the present study Cronbach's alpha for the NPAD–DLV was high (0.93). Other studies also found high values of Cronbach's alpha (range: 0.93–0.97) [6, 17, 19, 21, 22, 29, 34] indicating redundancy of items. Cronbach's alpha for the NDI–DLV in the present study (0.83) also falls within the range (0.74–0.92) reported by others [8, 17, 19, 20, 22, 32]. The variation in the Item–total score correlations for the NPAD–DLV and the NDI–DLV observed in the present study is similar with the variation found in other language versions (0.45–0.91 for the NPAD [6, 29, 34] and 0.45–0.84 for the NDI [20, 32]).

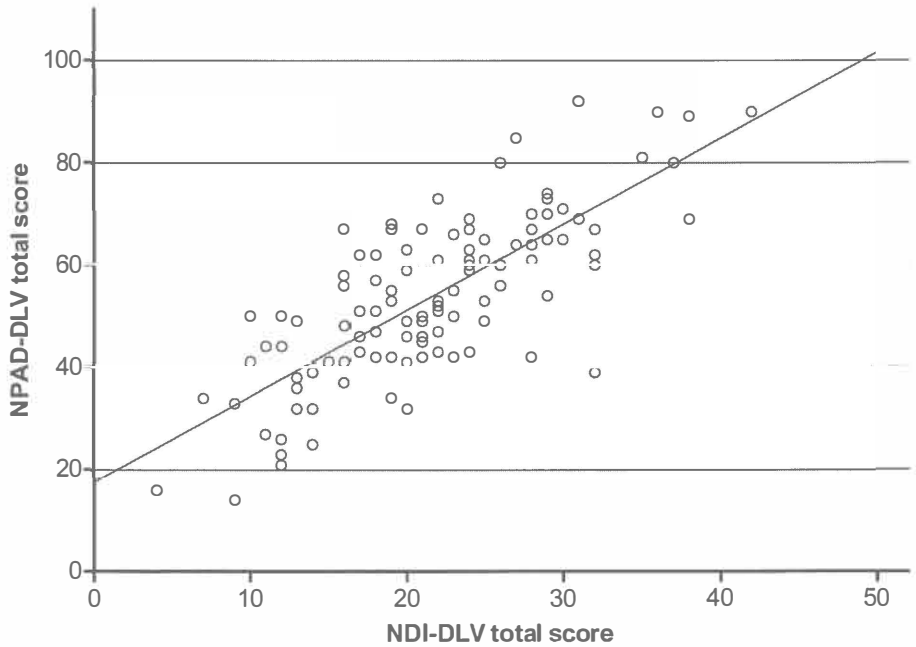


Figure 3.1 Scatterplot showing total scores of NDI–DLV and NPAD–DLV.

There is no established gold standard for assessment of neck pain disability. Therefore, criterion validity of the NPAD and NDI could not be analyzed [24]. To test the construct validity, comparisons were made with other constructs known to be associated with neck pain, neck pain related disability or generic health. The differences in the strength of the relationship between NPAD–DLV and NDI–DLV and all eight SF-36 domains with previous studies may be explained by differences such as patient setting, nature of neck condition, pain duration, and amount of neck pain related disability of the study samples [8, 20–22]. In the present study the correlation between the NPAD–DLV and VAS_{pain} was slightly higher than for the NDI–DLV and VAS_{pain} as hypothesized [17, 35]. The correlation of the NPAD–DLV and NDI–DLV with $VAS_{disability}$ in the present study was similar with that of other studies [17, 35]. The correlation between NPAD–DLV and NDI–DLV ($r = 0.77$) was similar with other studies (0.66–0.86), suggesting that these questionnaires measure comparable constructs [2, 10, 22, 35].

A potential limitation of this study was that the sample consisted largely of patients with moderate neck pain and disability. Although this may be expected in this tertiary rehabilitation setting, the validity of the NPAD–DLV and NDI–DLV should also be tested in general practice populations. Furthermore, the period between the baseline assessment and the second assessment was variable and the stability of VAS_{pain} and $VAS_{disability}$ between first and second assessment was assumed but not formally assessed. All our patients with CNP started rehabilitation after

completing the waiting period, indicating that their health status had not changed substantially [15]. Therefore, although we cannot be sure, this suggests that the potential impact of this weakness is unlikely to be substantial [15]. Finally, the hypotheses and the cut-off points that were used in the current study were based on previous studies without a methodically and qualitatively analysis of the validity of these studies.

A strength of this study is that to the author's knowledge for the first time a validity study is performed for the NPAD as well as the NDI in relation with SF-36 domain scores, VAS_{pain} and $VAS_{disability}$. Another strength is that the validity of the questionnaires is tested using explicit pre-defined hypotheses. The advantage of this method is its explicitness and transparency. Because the results are presented in detail, readers can develop and test their own hypotheses and perhaps interpret the same results differently. Further study with the NPAD–DLV is necessary to assess other measurement properties, such as responsiveness and minimally important change.

Conclusion

The NPAD–DLV and NDI–DLV are valid questionnaires to measure self-reported disability in patients with CNP within an outpatient tertiary rehabilitation setting.

Acknowledgements

The authors thank Marleen Speller for assistance in data collection.

REFERENCES

1. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, Sprangers MA, te Velde A, Verrips E (1998) Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol 51(11):1055–1068
2. Aslan E, Karaduman A, Yakut Y, Aras B, Simsek IE, Yagly N (2008) The cultural adaptation, reliability and validity of neck disability index in patients with neck pain: a Turkish version study. Spine (Phila Pa 1976) 33(11):E362–E365
3. Bicer A, Yazici A, Camdeviren H, Erdogan C (2004) Assessment of pain and disability in patients with chronic neck pain: reliability and construct validity of the Turkish version of the neck pain and disability scale. Disabil Rehabil 26(16):959–962
4. Bogduk N (1999) The neck. Baillieres Clin Rheumatol 13(2):261–285
5. Boonstra AM, Schiphorst Preuper HR, Reneman MF, Posthumus JB, Stewart RE (2008) Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain. Int J Rehabil Res 31(2):165–169

6. Bremerich FH, Grob D, Dvorak J, Mannion AF (2008) The neck pain and disability scale: cross-cultural adaptation into German and evaluation of its psychometric properties in chronic neck pain and C1-2 fusion patients. Spine 33(9):1018–1027

7. Clair D, Edmondston S, Allison G (2004) Variability in pain intensity, physical and psychological function in non-acute, nontraumatic neck pain. Physiother Res Int 9(1):43–54

8. Cook C, Richardson JK, Braga L, Menezes A, Soler X, Kume P, Zaninelli M, Socolows F, Pietrobon R (2006) Cross-cultural adaptation and validation of the Brazilian Portuguese version of the neck disability index and neck pain and disability scale. Spine 31(14):1621–1627

9. Crossley KM, Bennell KL, Cowan SM, Green S (2004) Analysis of outcome measures for persons with patellofemoral pain: which are reliable and valid? Arch Phys Med Rehabil 85(5):815–822

10. En MC, Clair DA, Edmondston SJ (2009) Validity of the Neck Disability Index and Neck Pain and Disability Scale for measuring disability associated with chronic, non-traumatic neck pain. Man Ther 14(4):433–438

11. Ferrari R, Russell AS (2003) Regional musculoskeletal conditions: neck pain. Best Pract Res Clin Rheumatol 17(1):57–70

12. Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT (1993) The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? BMJ 306(6890):1440–1444

13. Gay RE, Madson TJ, Cieslak KR (2007) Comparison of the neck disability index and the neck bournemouth questionnaire in a sample of patients with chronic uncomplicated neck pain. J Manipulative Physiol Ther 30(4):259–262

14. Hoving JL, O'Leary EF, Niere KR, Green S, Buchbinder R (2003) Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. Pain 102(3):273–281

15. Jorritsma W, de Vries GE, Geertzen JH, Dijkstra PU, Reneman MF (2010) Neck pain and disability scale and the neck disability index: reproducibility of the dutch language versions. Eur Spine J 19(10):1695–1701

16. Köke AJA, Heuts PHTG, Vlaeyen JWS (1996) Neck disability index. Pain knowledge center, university medical center maastricht, measurement instruments chronic pain. Pijn Kennis Centrum, Academisch Ziekenhuis Maastricht, Meetinstrumenten chronische pijn, Maastricht

17. Kose G, Hepguler S, Atamaz F, Oder G (2007) A comparison of four disability scales for Turkish patients with neck pain. J Rehabil Med 39(5):358–362

18. Landers MR, Cheung W, Miller D, Summons T, Wallmann HW, McWhorter JW, Druse T (2007) Workers' compensation and litigation status influence the functional outcome of patients with neck pain. Clin J Pain 23(8):676–682

19. Lee H, Nicholson LL, Adams RD, Maher CG, Halaki M, Bae SS (2006) Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. Spine 31(16):1841–1845

20. McCarthy MJ, Grevitt MP, Silcocks P, Hobbs G (2007) The reliability of the Vernon and Mior neck disability index, and its validity compared with the short form-36 health survey questionnaire. Eur Spine J 16(12):2111–2117

21. Monticone M, Baiardi P, Nido N, Righini C, Tomba A, Giovanazzi E (2008) Development of the Italian version of the neck pain and disability scale, NPDS-I: cross-cultural adaptation, reliability, and validity. Spine 33(13):E429–E434

22. Mousavi SJ, Parnianpour M, Montazeri A, Mehdian H, Karimi A, Abedi M, Ashtiani AA, Mobini B, Hadian MR (2007) Translation and validation study of the Iranian versions of the neck disability index and the neck pain and disability scale. Spine 32(26):E825–E831

23. Nieto R, Miro J, Huguet A (2008) Disability in subacute whiplash patients: usefulness of the neck disability index. Spine (Phila Pa 1976) 33(18):E630–E635

24. Offringa M, Assendelft WJJ, Scholten RJPM (2008) Introduction to evidence-based medicine (Inleiding in evidence-based medicine), Houten

25. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine 27(5):515–522

26. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC (2007) Minimal clinically important change of the neck disability index and the numerical rating scale for patients with neck pain. Spine (Phila Pa 1976) 32(26):3047–3051

27. Portney LG, Watkins MP (2000) Foundations of Clinical Research, 2nd Ed. Prentice Hall Health, Upper Saddle River, New Jersey

28. Riddle DL, Stratford PW (1998) Use of generic versus region-specific functional status measures on patients with cervical spine disorders. Phys Ther 78(9):951–963

29. Scherer M, Blozik E, Himmel W, Laptinskaya D, Kochen MM, Herrmann-Lingen C (2008) Psychometric properties of a German version of the neck pain and disability scale. Eur Spine J 17(7):922–929

30. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60(1):34–42

31. Trouli MN, Vernon HT, Kakavelakis KN, Antonopoulou MD, Paganas AN, Lionis CD (2008) Translation of the neck disability index and validation of the Greek version in a sample of neck pain patients. BMC Musculoskelet Disord 9:106

32. Vernon H, Mior S (1991) The neck disability index: a study of reliability and validity. J Manipulative Physiol Ther 14(7):409–415

33. Vos CJ, Verhagen AP, Koes BW (2006) Reliability and responsiveness of the Dutch version of the neck disability index in patients with acute neck pain in general practice. Eur Spine J 15(11):1729–1736

34. Wheeler AH, Goolkasian P, Baird AC, Darden BV (1999) Development of the neck pain and disability scale. Item analysis, face, and criterion-related validity. Spine 24(13):1290–1294

35. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel (2002) French translation and validation of 3 functional disability scales for neck pain. Arch Phys Med Rehabil 83(3):376–382

Chapter 4

Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index

Wim Jorritsma
Pieter U. Dijkstra
Grietje E. de Vries
Jan H. B. Geertzen
Michiel F. Reneman

Eur Spine J (2012) 21:2550–2557

ABSTRACT

Purpose: To investigate relevant change on the Neck Pain and Disability Scale (NPAD) and Neck Disability Index (NDI) and which questionnaire is the most responsive in patients with non-specific chronic neck pain (CNP).

Methods: Seventy-six patients with non-specific CNP in an outpatient tertiary rehabilitation setting were dichotomized into "improved" and "stable" based on global perceived effect (GPE) scores. To investigate relevant change minimal detectable change (MDC) and minimal important change (MIC) with the receiver operator characteristic (ROC) cut-off point were assessed. Comparison of responsiveness was performed using areas under the ROC curve (AUC) and correlations between change scores of NPAD and NDI, and GPE.

Results: MDC and MIC on NPAD (scale 0–100) were 31.7 and 11.5 points, respectively. MDC and MIC on NDI (scale 0–50) were 8.4 and 3.5 points, respectively. Changes should exceed this MDC or MIC cut-off to be interpreted as relevant. AUC was 0.75 for both NPAD and NDI. Correlations between change scores of NPAD and NDI, and GPE were, respectively, 0.48 (95% CI 0.29–0.64) and 0.49 (95% CI 0.30–0.64).

Conclusions: Relevant change on both NPAD and NDI assessed with MDC and MIC resulted in different cut-offs and consequently with different amounts of certainty that the patient is improved. Responsiveness of NPAD and NDI was similar.

INTRODUCTION

The most frequently used neck disability questionnaires are the Neck Pain and Disability Scale (NPAD) [37] and Neck Disability Index (NDI) [35], which are validated in several languages [4, 16, 21, 22, 38]. To evaluate the effect of treatment programs for neck disorders it is necessary that questionnaires are responsive, i.e., have the ability to detect clinical important changes over time. There is a need to define minimum changes in scores on questionnaires that are relevant from patients-, clinicians- or socioeconomic perspectives [34]. To determine relevant change two concepts of interpretability are described [1, 3, 8, 10, 34]. In a distribution-based approach the statistical characteristics of the sample are used to express the observed change in a standardized metric [8, 10, 34]. The most commonly used measure is the minimal detectable change (MDC) [3, 5–7, 9, 10, 16, 22, 29, 32, 36, 38, 40]. The MDC assesses the minimal magnitude of change required to be confident that the observed change reflects 'real' change and not measurement error [1, 8, 10, 30, 34]. A major limitation of distribution-based approaches is that they are statistical measures which by themselves do not provide a good indication of the clinical relevance of the observed change [8, 10, 34].

The anchor-based approach assesses which change on a questionnaire corresponds with an important change defined on an external criterion or anchor [for example global perceived effect (GPE)] [8, 10, 17, 34]. The most common method in this approach is the calculation of the minimal important change (MIC) determined by the receiver operator characteristic (ROC) curve cut-off point [6–8, 10, 23, 29, 33, 34, 40]. A major limitation of the anchor-based approach is the absence of a gold standard for the external criterion. A further limitation is that it does not take measurement precision into account and therefore does not necessarily imply statistical significance [8, 10, 34]. Hence, studies which apply both approaches are relevant for clinicians and researchers [8, 10, 34]. Moreover, there is a need of studies that assess relevant changes and compare the responsiveness of neck disability questionnaires applied at the same time to the same sample of patients using the same methods to investigate which questionnaire is most appropriate [28]. There are no studies assessing the MDC and the MIC as concepts of interpretability of relevant change for both NPAD and NDI. The aim of this study was to investigate relevant change on the NPAD and NDI and to investigate which questionnaire is most responsive in a single sample of patients with non-specific chronic neck pain (CNP) in an outpatient tertiary rehabilitation setting.

MATERIALS AND METHODS

Study sample

Patients with CNP were recruited from referrals from general practitioners or medical specialists for diagnostic procedures as well as advices and rehabilitation treatment in a tertiary university

center for rehabilitation in the Netherlands. To be admitted for a multidisciplinary pain rehabilitation, patients had to agree with the time-contingent approach to restore activities and to facilitate return to work. Inclusion criteria for this study were non-specific CNP (>3 months duration), admitted for outpatient rehabilitation, age between 18 and 65 years, and sufficient knowledge of the Dutch language to complete questionnaires. Neck pain was labeled as "non-specific" or mechanical when the neck pain was produced or aggravated by neck movements or sustained neck postures and no specific underlying pathology could be established [2, 13]. Exclusion criteria were status post neck surgery, co-morbidity severely diminishing physical or mental capacity, pregnancy, addiction to drugs, and extensive psychological or behavioral problems. Specific neck pain and exclusion criteria were assessed based on clinical examination with help of "red flags" and "orange flags" and based on the information of the referrals [25, 26, 31].

Procedures

Prior to the first visit (T0) a questionnaire to assess patient and clinical characteristics was filled out. During T0 a review of the medical history and a physical examination was performed. A second visit (T1) was scheduled, prior to the start of the multidisciplinary rehabilitation program. During T1 the patients filled out the NPAD and NDI. After completion of the program varying from 3 to 5 months (T2), patients filled out the NPAD, NDI, and the GPE. All patients signed informed consent for their data to be used for research. Data were gathered as part of care as usual between November 2006 and October 2010.

Measurements

The NPAD consists of 20 items [37]. Each item has a VAS of 100 mm with numeric anchors at 0, 1, 2, 3, 4, and 5 (each 20 mm apart). Item scores range from 0 (no pain or activity limitation) to 5 (as much pain as possible or maximal limitation). The total NPAD score ranges from 0 to 100 points. Higher scores indicate greater disability [37]. The NPAD has shown to be a reliable and valid measure of disability in different languages [4, 16, 18, 19, 21, 22, 38].

The NDI consists of ten items [35]. Each item has six different assertions expressing progressive levels of pain or limitation in activities. Item scores range from 0 (no pain or limitation) to 5 (as much pain as possible or maximal limitation). The total NDI score ranges from 0 to 5 points. Higher scores indicate greater disability [35]. The NDI has shown to be a reliable and valid measure of disability in different languages [6, 7, 18–22, 29, 36, 38, 40].

For the GPE patients were asked to rate their overall perception of change since beginning treatment ranging from 3 (completely recovered) to zero (no change) to -3 (worse than ever). The reliability of the GPE was moderate to good in patients with neck pain and chronic arthritis [14, 24] and the validity was fair to moderate in patients with neck pain [6, 7, 24, 29, 40].

Data analyses and interpretation

We dichotomized patients into two groups based on GPE scores. Patients were considered improved when they scored completely recovered (3) or much recovered (2) and stable when they scored slightly recovered (1) no change (0) or slightly worsened (-1). Baseline (T0 and T1) variables were compared between these groups using t tests for independent samples and Chi-square tests for categorical data.

Relevant change was analyzed by calculating the MDC and MIC. MDC was calculated as 1.96 x √2 x standard error of measurement (SEM). The SEM was calculated in stable patients as SD x √ (1 - r) where r is the test–retest reliability coefficient expressed in ICC value and SD is the standard deviation of the baseline scores [30, 34].

ROC curves were constructed to determine MIC for NPAD and NDI [11, 30]. The ROC cut-off point was calculated by identifying the point on the ROC curve nearest to the upper left-hand corner, which is considered to be the best cut-off for which the sum of the percentages of false positives and false negatives classifications ([1 - sensitivity] + [1 - specificity]) is smallest [11].

Responsiveness was assessed by examining areas under ROC-curve (AUC) and correlations between change scores of NPAD and NDI, and GPE. AUC was obtained to describe the ability of the NPAD and NDI to distinguish improved patients from stable patients [30]. AUC of 0.50 indicates the questionnaire has no diagnostic accuracy beyond chance, whereas a value of 1.00 would indicate perfect accuracy [30]. AUC of at least 0.70 was considered adequate [34].

A visual method called 'anchor-based MIC distribution' [11] method was used to integrate anchor-based and distribution-based approaches. For the improved and stable group the distribution of the change scores on the NPAD and NDI were depicted in a graph [11, 12]. All statistical analyses were performed with SPSS software, version 18.0. The critical value for significance was $p<0.05$.

RESULTS

During the recruitment period 391 patients with CNP were referred to the Center for Rehabilitation. A total of 129 patients, of which 4 were with status post neck surgery, were admitted for multidisciplinary outpatient rehabilitation. A total of 125 patients fulfilled inclusion criteria for this study. During the waiting period 14 patients decided not to start with the rehabilitation program because of practical reasons unrelated to the study. After the start of the rehabilitation program 35 patients decided not to continue because of lack of further interest or practical reasons. A dataset of 76 patients who completed the program was collected. The clinical characteristics of these patients and of the 35 dropouts are presented in Table 4.1. After the rehabilitation program 6 patients were completely recovered as assessed with GPE, 39 much recovered, 17 slightly recovered, 10 no change, 3 slightly worsened, 0 much worsened and 1

Table 4.1 Baseline characteristics of improved and stable patients, and dropouts

	Improved patients (n = 45)	Stable patients (n = 30)	Dropouts (n = 35)
Age (years)	37.7 (12.3)	39.5 (12.0)	39.2 (10.1)
Duration of chronic pain (months)	18.5 (9.3–58.5)[a]	24.0 (9.0–69.0)[a]	18.0 (7.5–48.0)[a]
Sick leave in the past year (weeks)	18.5 (19.4)	16.1 (17.6)	17.5 (20.6)
NDI (0–50)	21 (5.5)	21 (8.1)	23 (8.6)
NPAD (0–100)	50 (12.3)	53 (16.5)	56 (20.0)
VAS_{pain} (0–100)	52 (20.1)	52 (18.6)	54 (24.8)
Female, (%)	67	77	49
Pain radiating to, (%)			
Shoulder(s)	82	83	86
Upper arm(s)	51	40	54
Forearm(s)	33	20	46
Hand/fingers	27	17	40
Between shoulder blades	47	53	50
Pins and needles below elbow, (%)	36	33	38
Concomitant complaints, (%)			
Headache	84	63	71
Dizziness	36	27	35
Concentration problems	29	17	12
Nausea	11	13	12
Fatigue	69	53	62
Low back pain	33	31	49
Self-reported cause of neck pain, (%)			
Motor vehicle accident	56	40	37
Other trauma	9	13	17
Spontaneously/unknown	7	7	9
Stress	4	7	6
Work-related	9	10	14
Other	16	23	17
Previous treatment for neck pain, (%)	91	93	94
Education			
Low	7	0	0
Intermediate	69	79	82
High	24	21	18
Work status (self-employed/ employee), (%)	4/78	10/80	17/60
Involved in litigation, (%)	40	27	29

Values are means (SD) unless otherwise indicated. NPAD, Neck Pain and Disability Scale; NDI, Neck Disability Index; VAS, Visual Analog Scale.

[a] Median and interquartilerRange for duration of pain (months).

worse than ever. In total 45 (60%) patients were labeled as improved and 30 (40%) patients as stable. Baseline differences between improved and stable patients were non-significant, as were baseline differences between improved and stable patients on the one hand and dropouts on the other hand. More male patients dropped out than female patients.

The results for NPAD and NDI at baseline and follow-up and the change scores in the stable and improved groups are shown in Table 4.2. The ICC values calculated for stable patients were 0.52 (95% CI 0.33–0.67) for NPAD and 0.86 (95% CI 0.79–0.91) for NDI. The SEM values of the stable patients were 11.4 for NPAD and 3.0 for NDI. These values resulted in MDC values of 31.7 points for NPAD and 8.4 for NDI.

ROC curves for NPAD and NDI are presented in Figure 4.1. The ROC cut-off MIC was for NPAD 11.5 points (sensitivity 0.74; specificity 0.70) and for NDI 3.5 points (sensitivity 0.74; specificity 0.66). Changes should exceed these values of MDC and MIC cut-offs (31.7 and 11.5 for NPAD and 8.4 and 3.5 for NDI) to be interpreted as relevant. The AUC for NPAD was 0.75 (95% CI 0.62–0.87) and for NDI 0.75 (95% CI 0.64–0.87). The correlation between change scores of NPAD and NDI, and GPE were, respectively, 0.48 (95% CI 0.29–0.64) and 0.49 (95% CI 0.30–0.64).

The 'anchor-based MIC distribution' graphs for NPAD and NDI are presented in Figures 4.2a and 4.2b. These figures illustrate the effect of using MIC cut-off for change scores in the distribution of true and false positives and negatives. When the change score equals MIC, 26% of the anchor-based improved patients have a lower change score. They are considered false negatives because the sensitivity of NPAD and NDI = 0.74. When the change score equals MIC, 30% (NPAD) and 34% (NDI) of the anchor-based stable patients have higher change scores. They are considered false positives because specificity of NPAD = 0.70 and of NDI = 0.66.

Table 4.2 Baseline, follow-up and mean change scores of NPAD and NDI for the total, improved and stable group of patients

		Baseline Mean (SD)	Follow-up Mean (SD)	Change Mean (SD)	95% CI	p-value
NPAD	Total (n = 76)[a]	51 (14.0)	36 (19.0)	15 (17.4)	11.1 to 19.2	<.001
	Improved (n = 45)	50 (12.3)	29 (17.3)	21 (16.1)	15.7 to 25.7	<.001
	Stable (n = 30)	53 (16.5)	46 (16.9)	7 (16.2)	0.3 to 13.1	.04
NDI	Total (n = 76)[a]	21 (6.6)	15 (7.2)	6 (5.9)	4.2 to 7.0	<.001
	Improved (n = 45)	21 (5.5)	13 (6.2)	8 (6.3)	5.7 to 9.6	<.001
	Stable (n = 30)	21 (8.1)	18 (7.7)	3 (4.2)	1.2 to 4.4	.001

NPAD, Neck Pain and Disability Scale; NDI, Neck Disability Index.
[a] One patient scored 'worse than ever' and was not included in the improved or stable group, but was included in the total group.

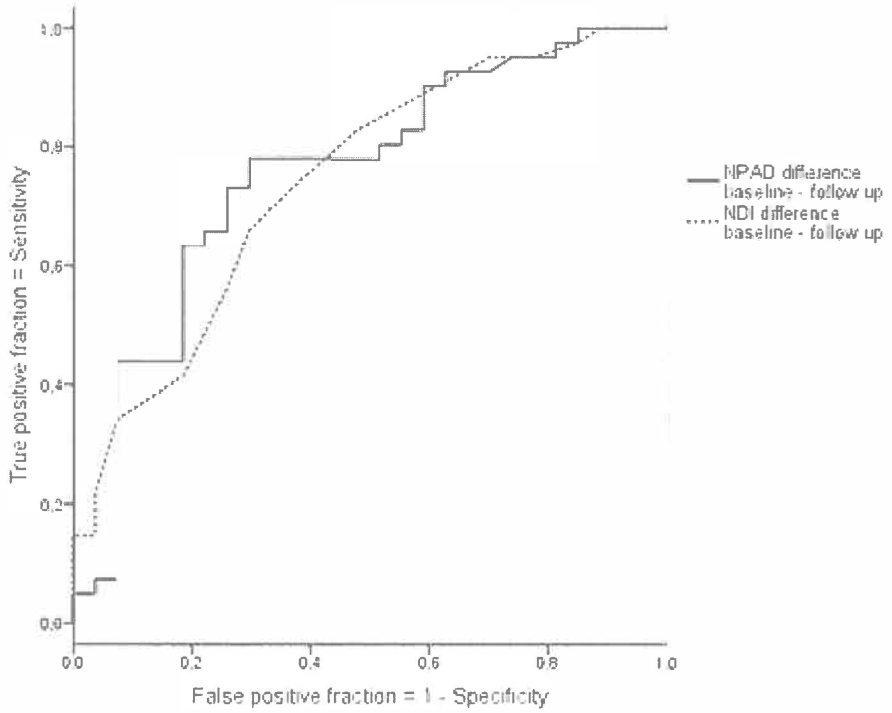


Figure 4.1 Receiver operator characteristic (ROC) curves of NPAD and NDI change scores.

DISCUSSION

This study demonstrated that relevant change on both NPAD and NDI assessed with MDC or MIC resulted in different cut-offs with different amounts of certainty that the patient is improved. Furthermore, it demonstrated that the responsiveness of NPAD and NDI was similar when using the AUCs and the correlations between change scores and the GPE.

There is no consensus regarding the number of SEMs required to express statistically clinically relevant change: 1 x SEM, 1.65 x SEM or 1.9 x SEM. We used the 1.96 x SEM to correspond with 95% CI. In the present study MDC for NPAD and NDI was 31.7 and 8.4, respectively. In a previous NPAD study [4] [mean baseline score 39.8 (SD 23.3)] the ICC was 0.97, the SEM 3.8 scale points, and follow up 1–2 weeks; therefore, the MDC of 10.5 was low compared with the present study. The ICCs of 0.52 for NPAD and 0.86 for NDI in the present study measured on "stable patients" was compared with the ICCs of 0.76 for NPAD and 0.84 for NDI in a previous study in the same setting with a retest interval of 18 days [18]. Larger instability of the NPAD may be explained by differences in operationalizations of neck disability between items of the NPAD and the NDI [35, 37]. Post hoc analysis showed that the amount of variation of the NPAD could be attributed

to significant differences in seven individual items (2, 6, 8–12) of the questionnaire. With an ICC of 0.76 for NPAD the MDC would be 22.4. Previous NDI studies report for MDC ranges between 1.7 and 13.4 [5–7, 29, 33, 36, 39, 40]. Apart from different patient populations, the observed

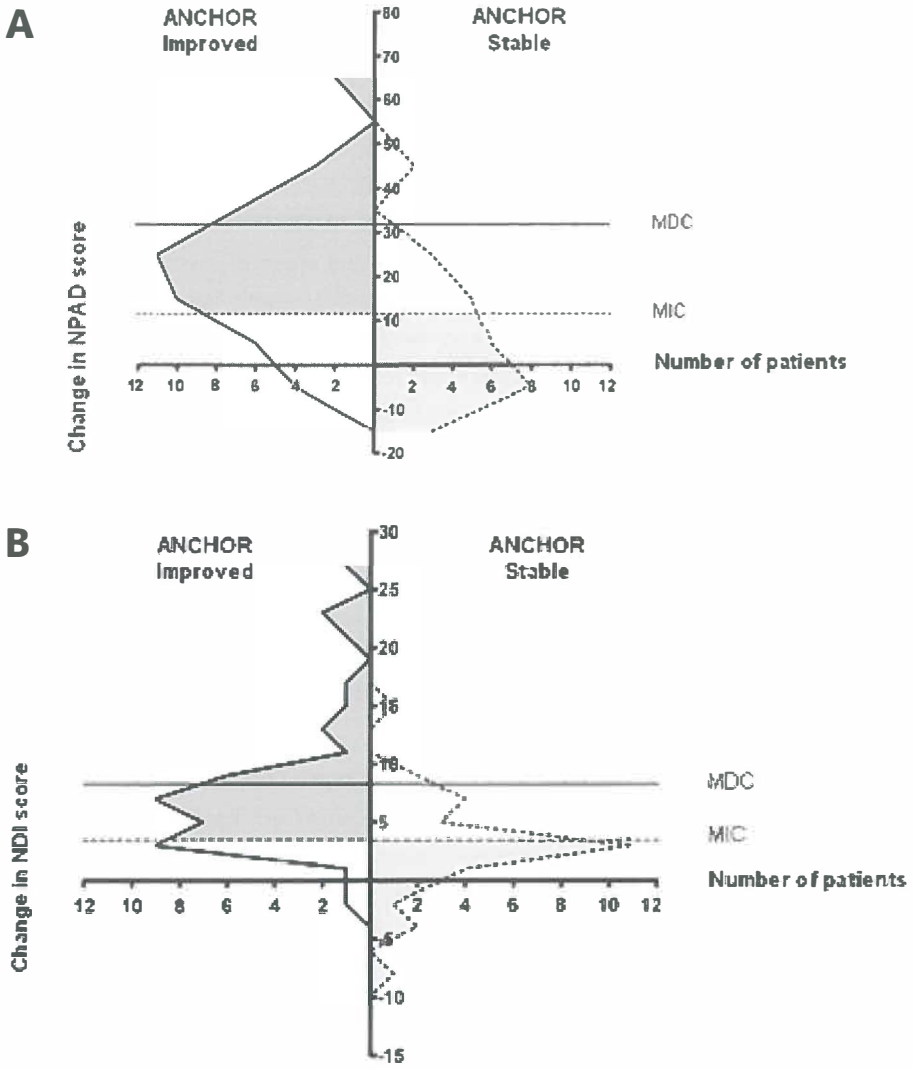


Figure 4.2 Distribution of NPAD-change scores in anchor-based improved and stable patients with indication of MIC at 11.5. At this point sensitivity = 0.74 and specificity = 0.70 **(a)**. Distribution of NDI-change scores in anchor-based improved and stable patients with indication of MIC at 3.5. At this point sensitivity = 0.74 and specificity = 0.66 **(b)**. Solid curve represents improved patients and dotted curve represents stable patients. The gray parts of the improved and stable patients represent the true positives (dark gray) and the true negatives (light gray), respectively. MDC is indicated at 31.7 for NPAD and at 8.4 for NDI.

differences are most likely the result of different formula for the MDC calculation (1.96 or 1.65 x √2 x SEM) and large ranges in SEM (0.60–4.4) in these NDI studies [5–7, 29, 33, 36, 39, 40].

MIC is defined as "the smallest change that is important to patients" [8, 10, 17, 33, 34]. How to classify the smallest important change and patients as improved or stable with GPE scale levels, is an arbitrary decision [5–7, 10, 29, 36, 38–40]. In most studies using GPE as external standard, a 15-point scale was used with ≥3 (moderately better) as cutoff to distinguish improved from stable patients [6, 7, 29, 39, 40]. We classified patients as improved when their score completely recovered or much recovered to reflect important improvement similar to other studies [11, 29]. Consequently, this may lead to overestimation of the MIC. In the present study the MICs for NPAD and NDI were 11.5 and 3.5. No values of MIC for NPAD have been reported by others. MIC for NDI has been reported to range from 3.5 to 9.5 [5–7, 29, 33, 39, 40]. Differences between these studies and the present study could be the result of several factors, such as different external criteria (prognostic estimate of change [33], Health Transition Item of SF-36 [5] and GPE by patient [6, 7, 29] or by patient and therapist [39]), the number of scale levels of the external criteria, the combination of scale levels to form the improved and stable group, characteristics of population (such as age, nature and acuity of neck condition, patient setting, baseline scores), treatment, and period of follow up [5–7, 29, 39, 40].

The AUC was used to determine the probability that the improved patient can be correctly distinguished from the stable patient. In this study NPAD and NDI both have an AUC of 0.75 which is a satisfactory result and in line with results found by other studies (range 0.57–0.90) [6, 7, 22, 29, 32, 33, 39, 40]. The AUC of 0.90 for the NDI was reported in a study using a prognostic estimate of change as external criterion made by clinicians at patient's initial visit [33]. In one study [22] responsiveness of NPAD and NDI was also compared using AUC. This study reported an AUC of 0.79 for both NPAD and NDI.

Clinicians should be aware of the fact that choosing either the MDC or the MIC cut-off gives different values and amounts of certainty on whether the observed change is relevant. Smaller values for the MIC were observed in almost all neck pain studies including the present study [5–7, 29, 39, 40]. Using the anchor-based MIC the proportion of false positives and false negatives is found to be the smallest. By raising the cut-off, the probability of false positives is reduced and the probability of false negatives is increased [12]. Applying the more conservative MDC, the certainty that the change score is relevant and larger than the measurement error, is high. The amount of certainty needed may depend on the consequences in patient care and could be a case by case decision.

For example for risk, full neck surgery or an expensive time-consuming multidisciplinary rehabilitation the more conservative MDC cut-off could be used, while in primary care setting the more liberal MIC cut-off could be used. On the other hand, socio-economic factors such as chance of returning to work as result of a therapy may be also of importance as external criterion for relevant change.

In the present study, the visual 'anchor-based MIC distribution' method was used whereby the distribution of the change scores on the NPAD and NDI was depicted in curves. The narrower the curves and the smaller the overlap of the curves, the smaller the chance of misclassification [12]. Both aspects of the curves largely depended on the correlation between change scores of NPAD and NDI and GPE as anchor [12]. In the present study these correlations were similar to those of most other NPAD (range 0.42–0.59) [16, 22, 38] and NDI (range 0.19–0.58) studies [6, 7, 22, 32, 38–40]. The GPE as external criterion to operationalize relevant change has been criticized because it consists of only one question and patient's ability to recall their previous health status is questionable [15, 27]. Any anchor-based approach is as good as the used external criterion and the methodology to define relevant change.

The present study is conducted in a university setting and is therefore representative of patients with CNP in a tertiary referral center. Percentage of females was similar to that of other responsiveness studies [4–7, 16, 21, 22, 29, 32, 33, 36, 38, 40]. Mean age in our study was lower (38.5 years) compared with other responsiveness studies [4–7, 16, 21, 22, 29, 32, 33, 36, 38, 40]. A potential limitation of this study is that the sample consisted largely of patients with moderate neck pain and disability. Although this may be expected in this tertiary rehabilitation setting generalizability beyond this setting cannot be assumed. The dropouts did not introduce bias because this study was aimed to measure the questionnaires and not the effect of the rehabilitation program. The strength of this study is that relevant changes were assessed with MDC and MIC on the NPAD and NDI and that a head-to-head comparison of the responsiveness of NPAD and NDI was performed. Further study of MDC, MIC, and responsiveness of NPAD and NDI is necessary to assess the measurement properties in other patient groups and also in comparison with other external criteria for relevant change.

Conclusion

Relevant change of both NPAD and NDI assessed with MDC and MIC resulted in different cut-offs with different amounts of certainty that the patient is improved. Responsiveness of NPAD and NDI was similar.

Acknowledgements

The authors thank Marleen Speller for assistance in data collection.

REFERENCES

1. Beaton DE (2000) Understanding the relevance of measured change through studies of responsiveness. Spine (Phila Pa 1976) 25:3192–3199
2. Bogduk N (1999) The neck. Baillieres Clin Rheumatol 13:261–285
3. Bolton JE (2004) Sensitivity and specificity of outcome measures in patients with neck pain: detecting clinically significant improvement. Spine (Phila Pa 1976) 29:2410–2417
4. Bremerich FH, Grob D, Dvorak J, Mannion AF (2008) The Neck Pain and Disability Scale: cross-cultural adaptation into German and evaluation of its psychometric properties in chronic neck pain and C1–2 fusion patients. Spine 33:1018–1027
5. Carreon LY, Glassman SD, Campbell MJ, Anderson PA (2010) Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J 10:469–474
6. Cleland JA, Fritz JM, Whitman JM, Palmer JA (2006) The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine (Phila Pa 1976) 31:598–602
7. Cleland JA, Childs JD, Whitman JM (2008) Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. Arch Phys Med Rehabil 89:69–74
8. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC (2007) Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 7:541–546
9. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY (2008) Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J 8:968–974
10. Crosby RD, Kolotkin RL, Williams GR (2003) Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 56:395–407
11. de Vet HC, Ostelo RW, Terwee CB, van der RN, Knol DL, Beckerman H, Boers M, Bouter LM (2007) Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 16:131–142
12. de Vet HC, Terluin B, Knol DL, Roorda LD, Mokkink LB, Ostelo RW, Hendriks EJ, Bouter LM, Terwee CB (2010) Three ways to quantify uncertainty in individually applied "minimally important change" values. J Clin Epidemiol 63:37–45
13. Ferrari R, Russell AS (2003) Regional musculoskeletal conditions: neck pain. Best Pract Res Clin Rheumatol 17:57–70
14. Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H (1999) Capturing the patient's view of change as a clinical outcome measure. JAMA 282:1157–1162
15. Gatchel RJ, Mayer TG (2010) Testing minimal clinically important difference: consensus or conundrum? Spine J 10:321–327

16. Goolkasian P, Wheeler AH, Gretz SS (2002) The neck pain and disability scale: test-retest reliability and construct validity. Clin J Pain 18:245–250

17. Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 10:407–415

18. Jorritsma W, de Vries GE, Geertzen JH, Dijkstra PU, Reneman MF (2010) Neck Pain and Disability Scale and the Neck Disability Index: reproducibility of the Dutch Language Versions. Eur Spine J 19:1695–1701

19. Jorritsma W, de Vries GE, Dijkstra PU, Geertzen JH, Reneman MF (2012) Neck Pain and Disability Scale and Neck Disability Index: validity of Dutch language versions. Eur Spine J 21:93–100

20. Köke AJA, Heuts PHTG, Vlaeyen JWS (1996) Neck Disability Index. Pain Knowledge Center, University Medical Center Maastricht, Measurement instruments chronic pain. (Pijn Kennis Centrum, Academisch Ziekenhuis Maastricht, Meetinstrumenten chronische pijn), Maastricht

21. Kose G, Hepguler S, Atamaz F, Oder G (2007) A comparison of four disability scales for Turkish patients with neck pain. J Rehabil Med 39:358–362

22. Lee H, Nicholson LL, Adams RD, Maher CG, Halaki M, Bae SS (2006) Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. Spine 31:1841–1845

23. Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 18(Suppl 3):374–379

24. Ngo T, Stupar M, Cote P, Boyle E, Shearer H (2010) A study of the test-retest reliability of the self-perceived general recovery and self-perceived change in neck pain questions in patients with recent whiplash-associated disorders. Eur Spine J 19:957–962

25. Nicholas MK, Linton SJ, Watson PJ, Main CJ, "Decade of the Flags" Working Group (2011) Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Phys Ther 91:737–753

26. Nordin M, Carragee EJ, Hogg-Johnson S, Weiner SS, Hurwitz EL, Peloso PM, Guzman J, van der Velde G, Carroll LJ, Holm LW, Cote P, Cassidy JD, Haldeman S (2009) Assessment of neck pain and its associated disorders: results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders. J Manip Physiol Ther 32:S117–S140

27. Norman GR, Stratford P, Regehr G (1997) Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 50:869–879

28. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine 27:515–522

29. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC (2007) Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine (Phila Pa 1976) 32:3047–3051

30. Portney LG, Watkins MP (2000) Foundations of Clinical Research, 2nd edn. Prentice Hall Health, Upper Saddle River

31. Rubinstein SM, van Tulder M (2008) A best-evidence review of diagnostic procedures for neck and low-back pain. Best Pract Res Clin Rheumatol 22:471–482

32. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M (2007) Responsiveness of pain and disability measures for chronic whiplash. Spine (Phila Pa 1976) 32:580–585

33. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B (1999) Using the Neck Disability Index to make decisions concerning individual patients. Physiother Can 51:107–112, 119

34. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34–42

35. Vernon H, Mior S (1991) The Neck Disability Index: a study of reliability and validity. J Manip Physiol Ther 14:409–415

36. Vos CJ, Verhagen AP, Koes BW (2006) Reliability and responsiveness of the Dutch version of the Neck Disability Index in patients with acute neck pain in general practice. Eur Spine J 15:1729–1736

37. Wheeler AH, Goolkasian P, Baird AC, Darden BV (1999) Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. Spine 24:1290–1294

38. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel M (2004) The ability to change of three questionnaires for neck pain. Jt Bone Spine 71:317–326

39. Young BA, Walker MJ, Strunce JB, Boyles RE, Whitman JM, Childs JD (2009) Responsiveness of the Neck Disability Index in patients with mechanical neck disorders. Spine J 9:802–808

40. Young IA, Cleland JA, Michener LA, Brown C (2010) Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. Am J Phys Med Rehabil 89:831–839

Chapter 5

The construct validity of the Short Form-36 Health Survey for patients with non-specific chronic neck pain

Wim Jorritsma
Michiel F. Reneman
Pieter U. Dijkstra
Grietje E. de Vries
Jan H. B. Geertzen

Submitted

ABSTRACT

Purpose: Self-reported disability related to neck pain can be measured with general health questionnaires. The validity of the Short Form-36 Health Survey (SF-36) in patients with non-specific chronic neck pain (CNP) in a tertiary outpatient rehabilitation setting is unknown. This study investigates construct validity of the SF-36 in these patients using 16 a priori formulated hypotheses.

Methods: Ninety-one patients admitted for rehabilitation completed the SF-36 prior to the rehabilitation program. SF-36 domain scores of patients with CNP were compared with general population reference values and standardized differences were calculated. For both the SF-36 physical and mental component summary (PCS and MCS) differences between primary and tertiary care setting, males and females, age groups, litigants and non-litigants, patients with and without compensation and with ≥3 versus ≤2 concomitant complaints were analyzed with independent t-tests. Difference between PCS and MCS scores was analyzed with a paired t-test.

Results: Twelve hypotheses were not rejected and 4 were rejected. All SF-36 domain scores were significantly lower than general population references values. The domain scores 'role physical', 'bodily pain', 'vitality', 'social functioning' and 'role emotional' were relevantly (≥1 SD) lower. SF-36-PCS and SF-36-MCS scores were significantly lower in tertiary care. SF-36-PCS score was significantly lower for patients with workers compensation, and patients with ≥3 concomitant complaints. SF-36-MCS score was significantly lower for the age group ≥39 years.

Conclusion: The SF-36 has good construct validity and can be used to measure self-reported general health in patients with non-specific CNP in outpatient tertiary rehabilitation.

INTRODUCTION

The prevalence of chronic neck pain (CNP) in the general population is about 20% [4, 22]. In the large majority of neck pain it is not possible to determine an underlying disease or which structure is affected. Hence the pain is labeled as "non- specific" or "mechanical" [3]. Self-reported disability in patients with neck pain is often measured with general health and region-specific questionnaires [23]. Unlike the region specific questionnaires the general health questionnaires cover broader emotional and social functions. The Short Form-36 Health Survey (SF-36) is a widely used general health questionnaire assessing health in 8 domains [14, 38]. Patients with neck pain report lower scores on all domains [8, 12, 15, 17, 19, 31] and thereby a poorer health than the general population [12, 31]. The acuity of neck pain, setting, age, percentage radiculopathy, percentage of litigation and workers compensation in the above mentioned studies were different. Litigation and workers compensation may play a negative role in perceived disability of patients with non-specific neck pain, chronic pain or back pain [2, 5, 8, 16, 26, 27, 32, 33].

Patients with non-specific CNP admitted for a multidisciplinary outpatient pain rehabilitation in a tertiary care setting are younger and report more disability, more frequently trauma as cause of neck pain, more concomitant complaints and higher percentages of litigation and workers compensation than patients with non-specific neck pain in primary care [10, 13, 25, 29, 35, 36].

Questionnaires should have good psychometric qualities, including different aspects of validity [23]. Construct validity is the degree to which the scores of an instrument are consistent with a priori hypotheses (e.g. with regard to relationships with similar or related constructs or differences between relevant groups) [18, 34]. The effect of non-specific CNP on the SF-36 health status in a tertiary outpatient rehabilitation setting is unknown. The aim of this study is to investigate the construct validity of the SF-36 in patients with non-specific CNP in this setting. Based on the guidelines described by the COSMIN group [18], a priori hypotheses were defined (Text Box 5.1) and outlined in Methods.

METHODS

Study sample

Patients with CNP were recruited from referrals by general practitioners or medical specialists for diagnostics, advice, and rehabilitation treatment in a tertiary university center for rehabilitation in The Netherlands. Inclusion criteria for this study were non-specific CNP (>3 months duration), admitted for outpatient rehabilitation, age between 18 and 65 years, and sufficient knowledge of the Dutch language to complete questionnaires. To be admitted for a multidisciplinary pain rehabilitation, patients had to agree with the time-contingent approach to restore activities and to facilitate return to work. Neck pain was labeled as non-specific or mechanical when

Text box 5.1 Sixteen a priory hypotheses for analyzing construct validity of the SF-36 in patients with chronic neck pain

The construct validity is not rejected when:

1. Physical domain scores 'role physical' (1a) and 'bodily pain' (1b) are relevantly (>1 SD) lower than general population reference values.
2. Mental domain scores 'vitality' (2a) and 'social functioning' (2b) are relevantly (>1 SD) lower than general population reference values.
3. PCS (3a) and MCS (3b) scores of patients with CNP in a tertiary rehabilitation setting are significantly lower than patients with neck pain in a primary care setting.
4. PCS scores are significantly lower than MCS scores.
5. Differences on PCS (5a) and MCS (5b) scores between males and females are not significant.
6. PCS (6a) and MCS (6b) scores of older (>39years) patients are significantly lower than younger (≤39 years) patients .
7. PCS (7a) and MCS (7b) scores of patients who are in litigation are significantly lower than patients who are not in litigation.
8. PCS (8a) and MCS (8b) scores of patients who are receiving workers compensation are significantly lower than patients who are not receiving workers compensation.
9. PCS scores of patients with 3 or more concomitant complaints are significant lower than patients with 2 or less concomitant complaints.

SF-36, Short Form-36 Health Survey; PCS, SF-36 Physical Component Summary; MCS, SF-36 Mental Component Summary.

the neck pain was produced or aggravated by neck movements or sustained postures and no specific underlying pathology could be discovered [3, 7]. Exclusion criteria were specific neck pain, status post surgery in the cervical region, comorbidity severely restricting physical capacity, pregnancy, addiction to drugs, and extensive psychological or behavioral problems. Specific neck pain and exclusion criteria were assessed based on clinical examination with help of 'red flags' and 'orange flags' and based on the information from the referrals [20, 21, 28].

Procedures

Prior to the first visit a questionnaire to assess patient and clinical characteristics was filled out. During the first visit a review of the medical history and a physical examination was performed. A second visit was scheduled, depending on the length of the waiting list and patient availability, 2–9 weeks after the first visit, but prior to the start of the outpatient rehabilitation program. During the second visit the patients filled out the SF-36. All patients signed informed consent for their data to be used for research. Data were gathered between November 2006 and September 2008 as part of care as usual. Ethics committee approval was not required because there was no deviation from care as usual.

Measurements

The SF-36 is a 36 item general health questionnaire [38]. It assesses health over the past 4 weeks in 8 domains: physical functioning (PF), role physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role emotional (RE), and mental health (MH). Scores for each domain range from 0 to 100, with higher scores indicating a better self-reported health. The physical component summary (PCS) and mental component summary (MCS) were constructed using weighted aggregates of all 8 domains of the SF-36 [37]. Higher PCS or MCS scores also imply better self-reported health. The SF-36 Dutch Language Version (DLV) has shown to be a reliable and valid instrument in community populations and in chronic disease populations [1].

Hypotheses

The 16 a priori hypotheses, listed in the Text Box, were based on previous studies [1, 2, 8, 9, 11, 12, 15, 17, 24, 26, 27, 31–33]. It was expected that the domain scores RP and BP (Hypotheses 1a and 1b) and VT and SF (Hypotheses 2a and 2b) were relevantly lower in patients with non-specific CNP than general population reference values [8, 12, 30, 31]. Significantly lower scores were expected on PCS (a) and MCS (b) in a Dutch tertiary rehabilitation setting than in patients with non-specific neck pain in a Dutch primary care setting (n = 146, mean PCS score 44.8 (SD 7.3) and mean MCS score 47.5 (SD 12.2)) [15, 17, 24] (Hypotheses 3a and 3b). Significant lower PCS scores than MCS scores were expected in patients with non-specific CNP (Hypothesis 4) [8, 9, 17, 27]. No significant differences between genders for PCS and MCS scores were expected (Hypotheses 5a and 5b) [17]. Significantly lower PCS (a) and MCS (b) scores were expected for patients of the study population above mean age (versus below mean age) (Hypotheses 6a and 6b) [1, 11, 17]. Significantly lower PCS (a) and MCS (b) scores were expected for patients who were in litigation, (versus no litigation) or who were receiving workers compensation (versus no workers compensation) because of their neck problems (Hypotheses 7a, 7b, 8a and 8b) [2, 8, 11, 17, 26, 27, 33]. Significantly lower PCS scores were expected in patients with 3 or more concomitant complaints than in patients with 2 or less concomitant complaints (Hypothesis 9).

Data analyses

Descriptive statistics were calculated for the SF-36 scores. Normality of the total scores of the SF-36 was analyzed with the Kolmogorov-Smirnov test. Differences between two groups (Hypotheses 3, 7-10) were analyzed with independent t-tests. Difference between PCS and MCS scores was analyzed with a paired t-test (Hypothesis 4). The construct validity was interpreted as good when at least 75% of the results corresponded with the hypotheses [34]. Normative data on the SF-36 were derived from a random nationwide sample (n = 1742; 56% men; mean

age: 47.6 year (SD 18.0)) of adults, 16 years or older [1]. We compared our SF-36 domain scores with the Dutch general population [1] and standardized differences were calculated as:

$$(Mean_{CNPsample} - Mean_{general\ population}) / Standard\ deviation_{general\ population}$$ [12].

To compare our SF-36 domain scores with other neck pain studies, standardized differences were also calculated for 2 other studies [8, 31]. All statistical analyses were performed with SPSS software, version 18.0. The critical values for significance were set at $p<0.05$.

RESULTS

In total 249 patients with CNP were referred to the Center for Rehabilitation during the time frame of data collection. From this group 101 were admitted for rehabilitation (41%). Three patients were excluded because of specific neck pain or status post cervical surgery. In total 98 patients were eligible for inclusion in this study. During the waiting period after the first visit 7 patients decided not to start with the rehabilitation program, so 91 patients with non-specific CNP were included in this study. The clinical characteristics of these patients are presented in Table 5.1. CNP was in 86% of the patients radiating into the shoulder(s), in 49% into the shoulder blades and in 47% into the arm(s). Frequently reported concomitant complaints were headache (77%) and fatigue (62%). The most frequent self-reported cause of neck pain was a motor vehicle accident. In total 57% of the patients was receiving working compensation and 33% was in litigation.

All SF-36 domain scores were significantly lower than general population norm values. The physical domain scores 'role physical' and 'bodily pain' were respectively 1.7 and 1.6 SD lower than general population norm values. The mental domain scores 'vitality' and 'social functioning' were respectively 1.4 and 1.3 SD lower than general population norm values (Table 5.2 and Figure 5.1).

The mean PCS score was 35.6 (SD±6.8) and the mean MCS score was 41.7 (SD±15.0). These scores were significantly lower than PCS and MCS scores in a Dutch primary care setting (mean difference = -9.20 (95% confidence interval (CI) -11.13 to -7.27), $p<0.05$ and mean difference = -5.80 (95% CI -9.39 to -2.21), $p<0.05$ respectively) [24]. Mean PCS scores were significantly lower than mean MCS scores (mean difference = -6.03 (95% CI -9.61 to -2.45), $p<0.01$).

Differences for the PCS and MCS scores between genders, age groups, litigation status, workers compensation status and concomitant complaints are presented in Table 5.3. PCS score was significantly lower for patients with workers compensation (versus non workers compensation) and patients with ≥3 concomitant complaints (versus ≤2 concomitant complaints). The mean number of concomitant complaints was 2.3 (SD 1.5). MCS score for the age group ≥39 years was significantly lower than for the age group <39 years. Hypotheses 6a, 7a, 7b and 8b were rejected.

Table 5.1 Clinical characteristics of the study sample (n = 91)

		Mean (SD) or Median (IQR)
Age (years)		38.7 (11.3)
Duration of chronic pain (months)[a]		18.0 (7.0 to 48.0)
Sick leave in the past year (weeks)		16.5 (18.6)
SF-36 PCS		35.6 (6.8)
SF-36 MCS		41.7 (15.0)
VAS_{pain} (scale 0–100)		57.4 (20.7)
	N	%
Female	60	66
Pain radiating to		
Shoulder(s)	78	86
Upper arm(s)	43	47
Forearm(s)	26	29
Hand/fingers	26	29
Between shoulder blades	43	49
Pins and needles below elbow	29	34
Concomitant complaints		
Headache	69	77
Dizziness	29	32
Concentration problems	16	18
Nausea	11	12
Fatigue	56	62
Low back pain	36	40
Self-reported cause of neck pain		
Motor vehicle accident	40	44
Other trauma	13	14
Spontaneously/unknown	6	7
Stress	4	4
Work-related	10	11
Other	18	20
Previously been treated for neck pain	82	91
Education		
Low	3	3
Intermediate vocational education	66	75
High	19	22
Work status (employed)	75	82
Workers compensation	52	57
Involved in litigation	29	33

[a] Median and interquartile range.
SF-36, Short Form-36 Health Survey; PCS, Physical Component Summary; MCS, Mental Component Summary; VAS, Visual Analog Scale.

Table 5.2 Means (SD) for SF-36 domains in patients with Chronic Neck Pain (CNP) and the general population

Domain	CNP (n = 83) Mean (SD)	General population[a] (n = 1742) Mean (SD)	Mean difference	95% CI			Standardized difference[b]
Physical Function	65.2 (18.7)	83.0 (22.8)	-17.8	-22.78	to	-12.82	-0.8
Role Physical	15.7 (29.1)	76.4 (36.3)	-60.7	-68.63	to	-52.77	-1.7
Bodily Pain	36.9 (17.3)	74.9 (23.4)	-38.0	-43.10	to	-32.90	-1.6
General Health	58.7 (21.1)	70.7 (20.7)	-12.0	-16.56	to	-7.44	-0.6
Vitality	41.0 (23.2)	68.6 (19.3)	-27.6	-31.89	to	-23.31	-1.4
Social Functioning	55.0 (24.9)	84.0 (22.4)	-29.0	-33.96	to	-24.04	-1.3
Role Emotional	53.7 (45.3)	82.3 (32.9)	-28.6	-35.99	to	-21.21	-0.9
Mental Health	65.3 (20.3)	76.8 (17.4)	-11.5	-15.36	to	-7.64	-0.7

[a] Aaronson et al., 1998.
[b] Standardized difference: $(Mean_{CNP\ sample} - Mean_{general\ population}) / SD_{general\ population}$.

Table 5.3 Results of independent t-tests for different groups in relation to PCS and MCS

		PCS		MCS	
		Mean (SD)	p-value	Mean (SD)	p-value
Age	≤39	35.6 (7.7)	.98	45.3 (14.1)	.02
	>39	35.6 (5.8)		37.6 (15.1)	
Gender	Male	34.5 (6.7)	.32	42.6 (15.5)	.69
	Female	36.1 (6.9)		41.2 (14.9)	
Litigation	Yes	33.7 (7.5)	.11	40.9 (14.0)	.70
	No	36.4 (6.3)		42.3 (15.2)	
WC	Yes	34.2 (6.6)	.02	40.5 (15.0)	.42
	No	37.6 (6.7)		43.2 (15.0)	
Concomitant complaints	≤2	37.2 (6.8)	.03	42.4 (15.6)	.76
	≥3	33.8 (6.3)		41.3 (14.7)	

PCS, Physical Component Summary; MCS, Mental Component Summary; WC, workers compensation.

DISCUSSION

Construct validity of the SF-36 is good, because 12 of 16 (75%) hypotheses were confirmed. To our knowledge this is the first report of the construct validity of the SF-36 in patients with non-specific CNP in a tertiary rehabilitation setting tested using a priori hypotheses. Thus, clinicians and researchers can use this questionnaire to assess self-reported general health in patients with non-specific CNP.

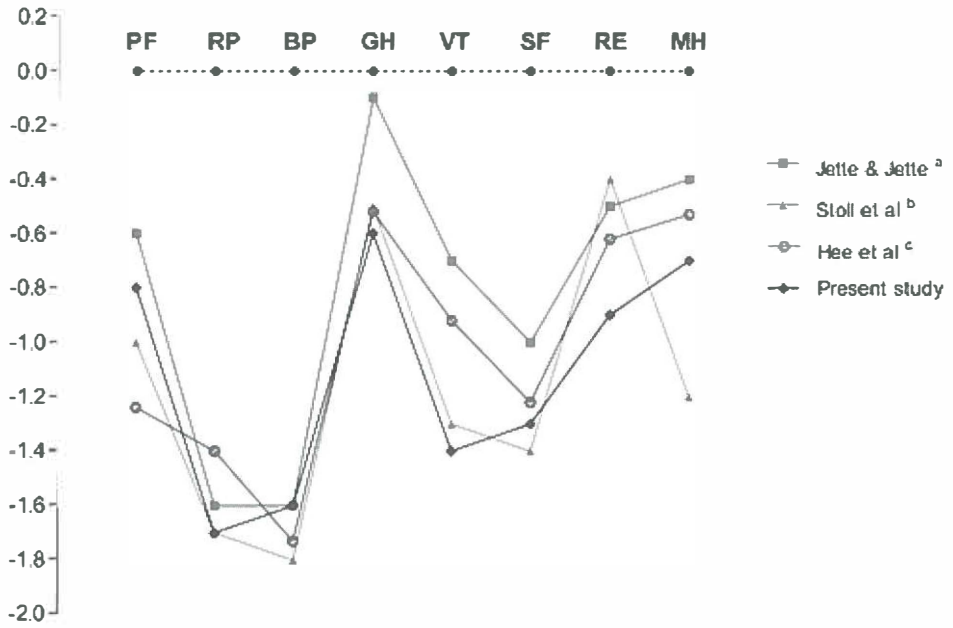


Figure 5.1 Comparison between standardized differences on SF-36 domains of patients with neck pain and general population norm values. The general population norm values are presented by the dots on the horizontal line at 0 on the y axis. PF, Physical Function; RP, Role Physical; BP, Bodily Pain; GH, General Health; VT, Vitality; SF, Social Functioning; RE, Role Emotional; MH, Mental Health. [a] Physical therapy outpatients; n = 358, 67% women; age: mean 42.0 (SD 13.4); pain duration: 29.7 (SD 23.5) days, acute: 19%, subacute: 53%, chronic: 28%; neck pain caused by: not presented. [b] Rehabilitation clinic inpatients (n = 98) and physical therapy outpatients (n = 42); n = 140, 64% women; age: mean 47.4 (SD 14.5); pain duration: not presented; neck pain caused by: muscular factors, degenerative changes of the cervical spine or whiplash injury. [c] Spine care centers patients; n = 2356, 53% women; age: mean 48.8 (SD 13.2); pain duration: acute: 14%, subacute: 9%, chronic: 70%; neck pain caused by: not presented.

In the current study 66% of patients was female which was similar to other studies (60 to 75%) [8, 9, 12, 17, 19, 27, 31]. Patients in the present study were somewhat younger (mean 38.7 years) than in other studies (ranging from 42.0 to 53.9 years) [8, 9, 12, 17, 19, 27, 31], which may limit generalization to older age groups in MCS score, but not in PCS score.

For the assessment of the validity of the SF-36 in patients with non-specific CNP, there is no established gold standard of general health to compare the SF-36 to. Therefore on the one hand comparisons with SF-36 domain scores in the general population and on the other hand comparisons of SF-36 PCS and MCS scores with values of patients with neck pain in a primary care setting were made. Overall the SF-36 domain scores in the present study make clear that

CNP has great impact on general health perception. This was in line with three other studies [8, 12, 31] with respectively a primary care setting (physical therapy outpatients), a mixture of rehabilitation clinic inpatients and physical therapy outpatients, and spine care center patients (Figure 5.1). The domain score profiles in the aforementioned studies and the present study are comparable. As expected, SF-36 domain scores of patients with neck pain in secondary/ tertiary care were overall lower than in primary care. The outpatient and inpatient rehabilitation groups appear more disabled and the lower VT, SF and MH scores of these patient groups may be explained by the presence of chronic pain and partly by the clinic inpatient setting.

As in other studies [8, 9, 17, 24, 27], the PCS score in the present study was lower than the MCS score. Apparently, in patients with neck pain physical health is affected to a greater extent than mental health. It is very likely that this is the result of the great impact of the PF, RP and BP on the PCS score. A further indication that physical health is effected to a greater extent is that higher correlations of PCS than of MCS with neck disability were reported [9, 17, 27]. As in a previous study, executed in a spinal tertiary referral setting we observed no statistically significant differences between genders for the SF-36 scores [17]. In the general population women scored significantly lower than man on all SF-36 domains [1, 11]. The difference in influence of gender on SF-36 domain scores between patients with CNP and general population samples may be explained by differences in work status, work place variables, education levels, litigation and socioeconomic status.

Because of the clinical characteristics of the patients in our tertiary rehabilitation setting we expected, as in patients with back pain, significantly lower PCS and MCS scores in the litigation group and in the workers compensation group [6, 8, 13]. Although the patients in the present study who were in litigation (33%) reported somewhat lower PCS and MCS scores than non-litigants, the differences were not significant. In a primary care setting 24% of patients with neck pain were in litigation and had significantly higher neck disability scores and significantly lower MCS scores than non-litigants [27]. Probably the contrast between groups with and without litigation is larger using a more specific neck pain related disability questionnaire than using the generic SF-36.

In the present study 57% of the patients were receiving workers compensation and their PCS and MCS scores were lower than of patients who were not receiving workers compensation; for the PCS scores this difference was significant as was reported in a primary care physiotherapy setting [27]. Because physical health in patients with neck pain is more effected than mental health, it is very likely that this has greater impact on PCS scores than on MCS scores in patients with workers compensation. In patients with neck pain (70% chronic) in spine care centers 7% of the patients were receiving workers compensation and their PCS and MCS scores were significantly lower than in patients without workers compensation [8]. However, differences were very small and clinically irrelevant, but due to a large sample size statistically significant [8].

There are limitations to consider in evaluating our research. First: the sample size is relatively small (with subgroups $n<50$), therefore our sample could have misestimated the subgroup

characteristics. However most of our a priori hypotheses based on previous neck pain studies were supported. Second: the sample consisted largely of patients with moderate neck pain related disability [13]. Therefore, the results of the present study cannot simply be generalized to other settings with other disability levels. Third: the hypotheses that were used in the current study were based on previous studies without a systematic appraisal of the quality of these studies.

A strength of this study is that the validity of the SF-36 in CNP patients is tested using explicit predefined hypotheses. The advantage of this method is its explicitness and transparency. The present study suggests that in the assessment of self-reported disability the use of SF-36 provides profound and comprehensive information of different aspects of physical and mental health in patients with non-specific CNP. Further research with the SF-36 is necessary to assess responsiveness in comparison with region specific questionnaires and socioeconomic outcomes in patients with neck pain in different settings.

Conclusion

The SF-36 has a good construct validity and can be used to measure self-reported general health in patients with non-specific CNP within an outpatient tertiary rehabilitation setting.

Acknowledgements

The authors want to thank Marleen Speller for assistance in data collection.

REFERENCES

1. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, Sprangers MA, te Velde A, Verrips E (1998) Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol 51:1055–1068
2. Blyth FM, March LM, Nicholas MK, Cousins MJ (2003) Chronic pain, work performance and litigation. Pain 103:41–47
3. Bogduk N (1999) The neck. Baillieres Clin Rheumatol 13:261–285
4. Bovim G, Schrader H, Sand T (1994) Neck pain in the general population. Spine (Phila Pa 1976) 19:1307–1309
5. Busse JW, Dufton JA, Kilian BC, Bhandari M (2004) The impact of non-injury-related factors on disability secondary to whiplash associated disorder type II: a retrospective file review. J Manipulative Physiol Ther 27:79–83
6. Deyo RA (2000) Pain and public policy. N Engl J Med 342:1211–1213
7. Ferrari R, Russell AS (2003) Regional musculoskeletal conditions: neck pain. Best Pract Res Clin Rheumatol 17:57–70

8. Hee HT, III Whitecloud TS,3rd, Myers L, Roesch W, Ricciardi JE (2002) Do worker's compensation patients with neck pain have lower SF-36 scores? Eur Spine J 11:375–381

9. Hermann KM, Reese CS (2001) Relationships among selected measures of impairment, functional limitation, and disability in patients with cervical spine disorders. Phys Ther 81:903–914

10. Hoving JL, de Vet HC, Koes BW, Mameren H, Deville WL, van der Windt DA, Assendelft WJ, Pool JJ, Scholten RJ, Korthals-de Bos IB, Bouter LM (2006) Manual therapy, physical therapy, or continued care by the general practitioner for patients with neck pain: long-term results from a pragmatic randomized clinical trial. Clin J Pain 22:370–377

11. Jenkinson C, Coulter A, Wright L (1993) Shortform 36 (SF36) health survey questionnaire: normative data for adults of working age. BMJ 306:1437–1440

12. Jette DU, Jette AM (1996) Physical therapy and health outcomes in patients with spinal impairments. Phys Ther 76:930–941

13. Jorritsma W, de Vries GE, Dijkstra PU, Geertzen JH, Reneman MF (2012) Neck Pain and Disability Scale and Neck Disability Index: validity of Dutch language versions. Eur Spine J 21:93–100

14. Keller SD, Ware JE,Jr, Bentler PM, Aaronson NK, Alonso J, Apolone G, Bjorner JB, Brazier J, Bullinger M, Kaasa S, Leplege A, Sullivan M, Gandek B (1998) Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 51:1179–1188

15. Klaber Moffett JA, Jackson DA, Richmond S, Hahn S, Coulton S, Farrin A, Manca A, Torgerson DJ (2005) Randomised trial of a brief physiotherapy intervention compared with usual physiotherapy for neck pain patients: outcomes and patients' preference. BMJ 330:75

16. Landers MR, Cheung W, Miller D, Summons T, Wallmann HW, McWhorter JW, Druse T (2007) Workers' compensation and litigation status influence the functional outcome of patients with neck pain. Clin J Pain 23:676–682

17. McCarthy MJ, Grevitt MP, Silcocks P, Hobbs G (2007) The reliability of the Vernon and Mior neck disability index, and its validity compared with the short form-36 health survey questionnaire. Eur Spine J 16:2111–2117

18. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. BMC Med Res Methodol 10:22

19. Monticone M, Baiardi P, Vanti C, Ferrari S, Nava T, Montironi C, Rocca B, Foti C, Teli M (2012) Chronic neck pain and treatment of cognitive and behavioural factors: results of a randomised controlled clinical trial. Eur Spine J 21:1558–1566

20. Nicholas MK, Linton SJ, Watson PJ, Main CJ, "Decade of the Flags" Working Group (2011) Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Phys Ther 91:737–753

21. Nordin M, Carragee EJ, Hogg-Johnson S, Weiner SS, Hurwitz EL, Peloso PM, Guzman J, van der Velde G, Carroll LJ, Holm LW, Cote P, Cassidy JD, Haldeman S (2009) Assessment of neck pain and its associated disorders: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. J Manipulative Physiol Ther 32:S117–40

22. Picavet HS, Schouten JS (2003) Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC(3)-study. Pain 102:167–178

23. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine 27:515–522

24. Pool JJ, Ostelo RW, Knol D, Bouter LM, de Vet HC (2010) Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain? Man Ther 15:111–116

25. Pool JJ, Ostelo RW, Knol DL, Vlaeyen JW, Bouter LM, de Vet HC (2010) Is a behavioral graded activity program more effective than manual therapy in patients with subacute neck pain? Results of a randomized clinical trial. Spine (Phila Pa 1976) 35:1017–1024

26. Rainville J, Sobel JB, Hartigan C, Wright A (1997) The effect of compensation involvement on the reporting of pain and disability by patients referred for rehabilitation of chronic low back pain. Spine (Phila Pa 1976) 22:2016–2024

27. Riddle DL, Stratford PW (1998) Use of generic versus region-specific functional status measures on patients with cervical spine disorders. Phys Ther 78:951–963

28. Rubinstein SM, van Tulder M (2008) A best-evidence review of diagnostic procedures for neck and low-back pain. Best Pract Res Clin Rheumatol 22:471–482

29. Schellingerhout JM, Heymans MW, Verhagen AP, Lewis M, de Vet HC, Koes BW (2010) Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. Spine (Phila Pa 1976) 35:E827–35

30. Simon GE, Revicki DA, Grothaus L, Vonkorff M (1998) SF-36 summary scores: are physical and mental health truly distinct? Med Care 36:567–572

31. Stoll T, Huber E, Bachmann S, Baumeler HR, Mariacher S, Rutz M, Schneider W, Spring H, Aeschlimann A, Stucki G, Steiner W (2004) Validity and sensitivity to change of the NASS questionnaire for patients with cervical spine disorders. Spine 29:2851–2855

32. Tait RC, Chibnall JT, Richardson WD (1990) Litigation and employment status: effects on patients with chronic pain. Pain 43:37–46

33. Teasell RW (2001) Compensation and chronic pain. Clin J Pain 17:S46–64

34. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34–42

35. Vonk F, Verhagen AP, Twisk JW, Koke AJ, Luiten MW, Koes BW (2009) Effectiveness of a behaviour graded activity program versus conventional exercise for chronic neck pain patients. Eur J Pain 13:533–541

36. Vos C, Verhagen A, Passchier J, Koes B (2007) Management of acute neck pain in general practice: a prospective study. Br J Gen Pract 57:23–28

37. Ware JE, Kosinski M, Keller SD (1994) SF-36 physical and mental health summary scales: a user's manual. The Health Institute, New England Medical Center, Boston

38. Ware JE, Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473–483

Chapter 6

Physical dysfunction and non organic signs in patients with chronic neck pain: explorative study into interobserver reliability and construct validity

Wim Jorritsma
Pieter U. Dijkstra
Grietje E. de Vries
Jan H. B. Geertzen
Michiel F. Reneman

Submitted

ABSTRACT

Study design: Repeated measurement design.

Objectives: To explore interobserver reliability of physical dysfunction severity (PDS) as measure for impairment of the cervical spine and modified cervical 'non organic signs' (mcNOS) as measure for behavioral signs, and to explore construct validity of PDS and mcNOS.

Background: PDS has been used for evaluation of treatment efficacy in controlled trials in primary care. cNOS was developed to assess abnormal illness behavior in patients with neck pain.

Methods: Two observers independently assessed PDS and mcNOS in 51 patients with chronic neck pain in an outpatient tertiary rehabilitation setting with a 3-week interval. Interobserver reliability for total scores of PDS and mcNOS was expressed as intraclass correlation (ICC). Interobserver agreement for each 'non organic sign' was calculated as absolute agreement and Cohen's kappa. Construct validity was expressed as Spearman correlation between PDS and mcNOS, with Neck Pain and Disability Scale (NPAD) and Numeric Rating Scale (NRS) for pain.

Results: Interobserver reliability for PDS and mcNOS was ICC=0.72 and ICC=0.78. Agreement for 'non organic signs' ranged from 63% to 88%. Kappa values ranged from 0.14 to 0.54. Correlations between PDS and mcNOS with NPAD were 0.26 and 0.49, and with NRS_{pain} 0.32 and 0.37.

Conclusion: Interobserver reliability of both PDS and mcNOS was acceptable. The interobserver agreement for the individual 'non organic signs' ranged from poor to acceptable. Construct validity of PDS and mcNOS appeared satisfactory.

INTRODUCTION

Prevalence of chronic neck pain (CNP) in the general population ranges from 10 to 21% [5, 30, 37]. In the large majority of cases it is not possible to determine the underlying cause of neck pain or the structure affected. Hence it is labeled as "non specific" or "mechanical" [4]. Disability is defined as an umbrella term for impairments, activity limitations and participation restrictions in a biopsychosocial model [49, 58]. Disability can be assessed clinically (based on clinical history and physical examination) during functional testing (for instance by a functional capacity evaluation (FCE)), by means of questionnaires, or by a combination of these methods [7, 10, 32]. An examination based on the assessment of cervical mobility, pain provocation with cervical movements [10, 22] and behavioral signs [28, 32] provides relevant information about a patient's clinical status. On average, little to fair association between physical impairment and behavioral signs with self-reported disability and pain has been reported [9, 20, 35, 42, 52, 56, 62]. On the individual patient's level, however, clinical management can be more effective when it is targeted at the problem areas identified during assessment of the clinical status. At the very minimum this assessment must have proof of reliability and validity.

The physical dysfunction severity (PDS) has been used for the evaluation of physical impairment and treatment efficacy in patients with neck pain [22, 26] and back pain [26] in controlled trials in primary care. PDS and perceived recovery were more sensitive to measure changes compared with self-reported neck pain related disability and neck pain intensity [22]. We modified PDS because we wanted to analyze pain report during movement by the patient (subscale 1) and the estimated limitations of movements by the examiner (subscale 2) separately. Because a reduced mobility of the shoulder girdle may be the result of a dysfunction of the cervicothoracic spine [34, 46], we also assessed the limitation in passive shoulder flexion (subscale 3). The PDS is a relatively simple instrument easily incorporated in daily practice. However, the clinimetric properties are still unknown.

Non organic signs (NOS) during physical examination were developed in 1980 [55] for patients with chronic back pain as screening tool to help identify patients who require more detailed psychosocial assessment. Although in a reappraisal of the interpretation of NOS [29] these signs were redefined as behavioral signs, the confusing description 'non organic signs' persisted [1, 2, 16]. Behavioral signs were described as 'responses to physical examination effected by fear in the context of recovery from injury and the development of chronic incapacity' [29]. In a structured evidence-based review [16] it was disputed that NOS does not discriminate 'organic' from 'non organic' problems. However most patients both have a physical problem and varying degrees of behavioral signs [29]. Consequently, isolated behavioral signs should not be overinterpreted. Moreover behavioral signs in patients with non-specific and specific chronic back pain were separable from and independent from the standard physical findings of non-specific and specific lumbar spine disorders [55]. Another point of criticism was that widespread tenderness may represent an organic phenomenon [16]. However without for

example a great amount of acute inflammation or signs of a systemic disease, widespread tenderness and in general hypersensitivity to mechanical, thermal or chemical stimulation are common features in chronic pain [15, 44, 47]. It has been suggested that central sensitization and changes in endogenous descending pain modulation mechanisms operating at somatic, cognitive, emotional and behavioral levels are the underlying mechanism of hypersensitivity in chronic pain syndromes [15]. Hence it is reasonable to assume that *a patient's* psychosocial status influences behavioral responses (such as facilitated pain responses with movements or manual pressure and tense muscles with movements) during physical examination in patients with chronic pain. Fishbain et al. [16] also reported inconsistent evidence in that NOS can be assessed reliably and that NOS does not correlate with psychological distress. A recent study however reported a moderate interobserver and a good intraobserver reliability of NOS in patients with chronic low back pain in an outpatient rehabilitation center [1]. Moreover the NOS score was positively associated with pain intensity, low back mobility, sick leave and higher scores for depression, psychological distress and psychopathology [2].

In 2000 a standardized set of cervical NOS (cNOS) was developed as a tool to assess abnormal illness behavior in patients with neck pain [45]. Physical examination signs of cNOS were classified into five categories based on lumbar NOS. The tests in the categories *tenderness*, *regional disturbances* and *overreaction* were extrapolated to the cervical spine [45]. Two additional tests for cNOS were specifically developed in the categories *simulation* and *distraction* [45]. In the simulation category the patient should have the idea that the neck area is being tested while in reality it is not. In fact the cNOS simulation test (the standing rotation of head/shoulders/trunk/pelvic as one unit over the pelvis by the examiner) is a simulation of low back rotation [53]. Therefore we transformed this test into a modified cNOS (mcNOS). In 'cervical rotation test' of cNOS the criterion for a positive behavioral sign was that the maximal active rotation while seated should be less than 50% of normal in each direction. In our opinion the 'cervical rotation test' in the cNOS is not an adequate distraction test. As in the lumbar NOS distraction test (straight leg raising in lying supine versus 'flip test' in sitting) the observations during the routine assessment should be checked for inconsistencies in another test situation [55]. As an additional consequence we also changed this test in the mcNOS. Kappas of interobserver agreement for the individual cervical 'non organic signs' ranged from 0.08 to 1.00 [45], but prevalence of positive signs just like interobserver reliability for total scores of cNOS were not reported. cNOS was associated with prolonged disability [28] and worker's compensation status [19].

The primary aim of this study was to explore interobserver reliability of PDS and mcNOS in patients with CNP in an outpatient tertiary rehabilitation setting. The secondary aim was to explore construct validity of PDS and mcNOS by testing 5 a priori hypotheses based on previous studies about the expected relationships of PDS and mcNOS scores with self-reported disability assessed with the Neck Pain and Disability Scale (NPAD) and pain intensity assessed with the Numeric Rating Scale (NRS) [6, 11, 24, 39, 57, 60, 61].

We expected positive associations between PDS scores and respectively NPAD scores (Hypothesis 1) and NRS_{pain} scores (Hypothesis 2) [9, 20, 42, 62]. We also expected positive associations between mcNOS scores and respectively NPAD scores (Hypothesis 3) and NRS_{pain} scores (Hypothesis 4) [35, 56]. The mcNOS was expected to correlate stronger with NPAD than PDS with NPAD (Hypothesis 5) [35, 56]. Basing ourselves on the abovementioned studies we expected to observe little to fair correlations in hypotheses 1 and 2 and fair correlations in hypotheses 3 and 4.

METHODS

Study sample

Patients with CNP were recruited for rehabilitation in a Center for Rehabilitation in The Netherlands from referrals by general practitioners or medical specialists. Inclusion criteria for this study were: non-specific CNP (>3 months duration), admitted for outpatient rehabilitation, age between 18 and 65 years, less than 2 years out of work due to CNP or still at work with frequent sick leave (on average >5% in the last 2 years) due to neck pain, and sufficient knowledge of the Dutch language to complete questionnaires. Exclusion criteria were: status post neck surgery, co-morbidity diminishing physical capacity, co-morbidity with severe negative consequences for physical and/or mental functioning (e.g. psychiatric disease), pregnancy, addiction to drugs.

Procedures

Prior to the first visit a questionnaire to assess patient and clinical characteristics was filled out. During the first visit a clinical observation was performed including the assessment of Numeric Rating Scale $(NRS)_{pain}$, PDS and mcNOS. Immediately afterwards patients filled out the Neck Pain and Disability Scale (NPAD). As part of care as usual a second visit was scheduled depending on patient's availability, 1–5 weeks later, but prior to the start of the rehabilitation program. During this visit a second examiner assessed PDS and mcNOS and in addition a neck FCE was performed [41]. The second examiner and patients were blinded for results of the first examination. All patients signed informed consent for their data to be used for research. As usual data were gathered as part of the care between November 2006 and May 2009. Ethics committee approval was not required because there was no deviation from usual care.

Measurements

Physical dysfunction severity (PDS)

We modified and operationalized PDS into 3 subscales: 1. Pain during active cervical movement, 2. Limitation of passive cervical range of movement (ROM) and 3. Limitation of passive shoulder flexion ROM. During the examination patients were sitting actively on a chair or examination table.

For the assessment of subscale 1, patients were asked to move their head and neck as far as possible in the following directions: flexion, extension, lateral flexion to the left and right, and rotation to the left and right, while avoiding compensatory spinal movements and to report presence of pain (yes or no). Subscale 1 ranged from 0 (all neck movements without pain) to 6 (all neck movements with pain).

For the assessment of subscale 2 patients were asked to perform active flexion, extension, lateral flexion to the left and right and rotation to the left and right of the cervical spine while the examiner gently applied pressure to guide the movement to the end of ROM. The examiner assessed the movements visually and labeled them as not limited = 0, slightly = 1 (5–24%), moderately = 2 (25–49%) or severely = 3 (≥50%) limited. The values of the age group 20–30 (flexion 60–70˚, extension 60–70˚, lateral flexion 40–45˚ and rotation 80–90˚) were used as 'gold standards' for all subjects in the study to estimate relative 'limitations' in cervical ROM [8, 14, 27, 31]. Subscale 2 ranged from 0 (all neck movements without limitation) to 18 (all neck movements with severe limitation). For the assessment of subscale 3 patients were asked to perform active shoulder flexion, while the examiner gently applied pressure to guide the movement to the end of ROM. The examiner assessed the movements visually and as in subscale 2 labeled them as not limited = 0, slightly = 1, moderately = 2 or severely = 3 limited. For the passive shoulder flexion a reference value of 170–180˚ was used [36]. Subscale 3 ranged from 0 (both shoulders without limitation) to 6 (both shoulders with severe limitation). The total score of the PDS ranged from 0 (no dysfunction) to 30 (extremely severe dysfunction). Criteria for scoring of the PDS are described in Appendix 6.1.

Modified cervical non organic signs (mcNOS)

In the present study a standardized set of 13 physical examination signs classified into five categories is used. Criteria for test interpretation are described in the Appendix 6.2. The set is based on the categories *tenderness*, *regional disturbances* and *overreaction* of the lumbar NOS and cNOS [45, 55]. The cNOS was modified in the categories *simulation* and *distraction* [45]. For the reason described in the introduction we developed a new *simulation* rotation test. In this test the patient is sitting, hands folded behind the neck and the elbows in front almost together. The examiner rotates the patient's trunk to the right and left using the patient's shoulders, while no rotation in the cervical spine occurs. The new test may have greater relevance as simulation

test because the hands of the patient are on the painful neck region, the apparent rotation of the head of the patient and the focus of the examiner on the neck region by asking about neck pain. In the new *simulation* 'axial loading' test the patient is sitting and the examiner applies manually a few pounds of pressure on the shoulders of the patient. This test is comparable with the axial loading test of the NOS. For the reason described in the introduction we developed new *distraction* tests. Because extension and rotations are the most frequently reported limitations, we designed extension and rotation distraction tests [27]. In the *distraction* extension test the patient lies prone supported on flexed elbows and is asked to look forward to a point high on the wall. In this test a maximal extension of the neck will be reached and the outcome can be used in the determination of the patient's consistency of presentation by comparing ROM with seated maximal, active extension. In the *distraction* rotation (left/right) test the patient lies prone with the head rotated to one side and is asked to lift the ipsilateral in elbow extended arm out of 90 degrees abduction position in the shoulder. During this functional test the neck will be rotated maximally. When during routine physical examination a remarkable limitation of the seated maximal active rotation is demonstrated, this finding can be checked on consistency with the distraction rotation test. The total score of the mcNOS ranges from 0 (no signs positive) to 13 (all signs positive).

Neck pain and disability scale (NPAD)

The NPAD consists of 20 items [57]. Each item has a visual analog scale of 100 mm with numeric anchors at 0, 1, 2, 3, 4 and 5 (each 20 mm apart). Item scores range from 0 (no pain or activity limitation) to 5 (as much pain as possible or maximal limitation). The total NPAD score ranges from 0 to 100 points. Higher scores indicate greater disability [17, 57]. The NPAD has shown to be a reliable and valid measure of disability in different languages [6, 24, 61].

Numeric rating scale for pain (NRS_{pain})

The NRS_{pain} is an 11-point rating scale in which 0 = no pain and 10 = worst pain imaginable. Patients were asked to rate their current pain at the start of the physical examination. The NRS_{pain} has shown to be a reliable and valid measure [11, 39, 60].

Data analysis and interpretation

A sample size of at least 50 patients is considered adequate for the assessment of reliability and construct validity [50]. Allowing for an attrition rate of 1 in 6 we needed to recruit 60 patients. For each examiner the mean, standard deviation (SD) and range were calculated for total scores of PDS and mcNOS. To calculate the prevalence (expressed as percentage) of each cervical 'non organic sign' the number of positive observations was defined as the sum for each sign found by both examiners divided by 2.

Interobserver reliability for the total scores of the PDS and mcNOS was expressed as intraclass correlation (ICC(2,k)). Correlations were interpreted as follows: $0.75 \leq r \leq 1.0$ as good, $0.50 \leq r \leq 0.75$ moderate, $0.25 \leq r \leq 0.50$ fair and $0.00 \leq r \leq 0.25$ little or none [40]. ICCs of 0.75 or higher were interpreted as acceptable reliability. Interobserver reliability of PDS subscales was determined by calculation of the squared weighted kappa. Interobserver agreement of each 'non organic sign' was determined by calculation of percentage of absolute agreement and Cohen's kappa. A kappa value ≥0.40 was a criterion for an acceptable reliability [13, 48]. An absolute agreement of ≥80% was also a criterion for an acceptable agreement. All statistical analyses were performed with SPSS software, version 18.0. Squared weighted Cohen's kappas were calculated with AGREE. The critical values for significance were set at $p < 0.05$.

RESULTS

During the recruitment period 62 patients were admitted for outpatient rehabilitation. A total of 60 patients fulfilled inclusion criteria for this study. During the waiting period after the first visit, 9 patients decided not to start with the rehabilitation program, because of practical reasons unrelated to the study. A complete dataset of 51 patients was collected. The clinical characteristics of these patients (mean age 38.5±10.9 years, 63% female) are presented in Table 6.1. The mean interval between observations was 20 days (SD 7.4).

The ICC for the PDS was 0.72 (95% CI 0.50 to 0.85) and for the mcNOS 0.78 (95% CI 0.61 to 0.88) (Table 6.2). The percent agreement ranged from 63% for deep tenderness to 88% for distracted extension and right rotation. The kappa coefficient ranged from 0.14 for distracted left rotation to 0.54 for rubbing/clutching the affected area (Table 6.3). All 5 hypotheses with respect to the construct validity of PDS and mcNOS were supported (Table 6.4).

DISCUSSION

In the present study interobserver reliability based on ICCs of PDS and mcNOS was acceptable. The interobserver agreement for the individual 'non organic signs' ranged from poor to acceptable. Construct validity of PDS and mcNOS appeared satisfactory and in agreement with expected associations between PDS and mcNOS with NPAD and NRS_{pain}.

Observer differences

Small but statistically significant differences were observed between the first and second examiner in PDS and mcNOS. One might consider several explanations for these differences in outcome: systematic and random differences. The first examination may induce a change in patient behavior, anticipation or a change in the presence of muscle tension. Pain provoked

Table 6.1 Clinical characteristics of the study sample at baseline (n = 51)

	Mean (SD) or Median (IQR)	Min–Max
Age (years)	38.5 (10.9)	21–57
Duration of chronic pain (months)	14.0 (6.5 to 24.5)[a]	4–276
Sick leave in the past year (weeks)	22 (17.9)	0–52
NPAD (0–100)	52 (15.2)	12–92
NRS_{pain} (0–10)	4.0 (2.7)	0–8
	N	%
Female	32	63
Pain radiating to		
Shoulder(s)	44	86
Upper arm(s)	24	47
Forearm(s)	16	31
Hand/fingers	14	28
Between shoulder blades	20	39
Pins and needles below elbow[b]	17	35
Concomitant complaints		
Headache	39	77
Dizziness	21	41
Concentration problems	13	26
Nausea	7	14
Fatigue	32	63
Low back pain	19	38
Self reported cause of neck pain		
Motor vehicle accident	27	53
Other trauma	5	10
Spontaneously/unknown	2	4
Stress	2	4
Work-related	4	8
Other	11	22
Education[c]		
Low	2	4
Intermediate vocational	37	74
High	11	22
Work status (self-employed/employee)	5/46	10/90
Workers compensation	37	73
Litigation	20	40

SD, standard deviation; IQR, interquartile range; NPAD, neck pain and disability scale; NRS, numeric rating scale.
[a] Median and interquartile range; [b] 2 missing data; [c] 1 missing.

Table 6.2 Mean scores, mean differences and reliability between Examiner 1 and 2 on mcNOS, PDS and PDS subscales

	Examiner 1		Examiner 2		Difference		p-value	ICC (95%CI)
Scale (scoring range)	Mean	SD	Mean	SD	Mean	SD		
mcNOS (0–13)	2.7	2.4	3.4	2.7	-0.7	2.1	.03	0.78 (0.61 to 0.88)
PDS (0–30)	6.8	3.7	8.1	5.3	-1.4	4.1	.02	0.72 (0.50 to 0.85)
								Kappa
PDS Subscale 1 (0–6)	4.1	2.1	3.9	2.2	0.2	2.2	.47	0.50[a]
PDS Subscale 2 (0–18)	2.4	2.7	4.0	3.9	-1.6	3.1	<.01	[b]
PDS Subscale 3 (0–6)	0.2	0.5	0.2	0.6	-0.0	0.5	.57	0.52[a]

mcNOS, modified cervical non organic signs; PDS, physical dysfunction severity; ICC, intraclass correlation coefficient; SD, standard deviation.
Subscale 1, pain during active cervical movement; Subscale 2, limitation of passive cervical range of movement; Subscale 3, limitation of passive shoulder flexion range of movement.
[a] squared weighted. [b] Kappa could not be calculated because of lack of variation in cell filling.

during the first examination may have influenced the outcome of tests during the second examination. The physical examinations were carried out by two experienced clinicians. However, differences in techniques between the examiners, differences in interpretation of the same signs despite extensive standardization of the test performance and interpretation prior to the study, and the different settings of the examinations (1st session in the hospital, and 2nd

Table 6.3 Prevalence, absolute agreement (%) and Cohen's kappa coefficients between Examiner 1 and 2 for each sign of the mcNOS

	Examiner 1 % positive observations	Examiner 2 % positive observations	Prevalence (%)	Absolute agreement (%)	Cohen's kappa
I Tenderness					
1. Superficial	16	19	18	81	.36
2. Deep	43	29	36	63	.22
II Simulation					
3. Axial loading	26	35	31	71	.33
4. Simulated rotation	22	35	29	65	.15
III Distraction					
5. Extension	10	6	8	88	.19
6. R-Rotation	8	12	10	88	.33
7. L-Rotation	10	10	10	85	.14
IV Regional disturbances					
8. Motor loss	4	15	10	74	.19
9. Sensory loss	4	30	17	87	.17
V Overreaction					
10. Disproportional verbalization	49	42	46	75	.50
11. Grimacing due to pain	41	42	42	67	.31
12. Rubbing/clutching the affected area	20	21	21	85	.54
13. Stiff, rigid or slow movement	18	48	33	69	.36

mcNOS, modified cervical non organic signs.

Table 6.4 Spearman correlations between PDS, mcNOS, NPAD and NRS_{pain}

	NPAD	NRS_{pain}	mcNOS
PDS	.26	.32[a]	.39[b]
mcNOS	.49[b]	.37[b]	-

PDS, physical dysfunctioning severity; mcNOS, modified cervical non organic signs; NPAD, neck pain and disability scale; NRS, numeric rating scale.
[a] <0.05; [b] <0.01.

in a testing laboratory after averagely 20 days) may have influenced the outcomes. It might have been possible that changes in clinical status occurred between the sessions due to the effect of a waiting period before the start of the rehabilitation program, due to the effect that patients were possibly tested at different times of the day and random fluctuations in CNP that may have occurred.

Physical dysfunction severity

To date it is unclear which physical test(s) should be used in clinical practice to assess actual physical impairment of the cervical spine but relatively simple and quick methods are preferred [3]. Because pain and limitation of movements are the most frequently mentioned complaints in patients with neck pain it was relevant to integrate both in respectively subscale 1 and 2 of the PDS. Subscale 2 consisted of passive ROMs because these were more reliable and provided more objective information than active ROM measurements [14]. The Task Force on Neck Pain and its Associated Disorders reported that visual estimation of active ROM by clinicians was as reliable as using an external device while others reported that quantification of passive and active neck ROM was more reliable when using a device [12, 33, 59].

Because of the observed differences on the PDS we analyzed its subscales, whereby only subscale 2 showed significant differences. The interobserver reliability on PDS-subscale 1 was similar compared to a previous study [10]. The interobserver reliability (kappa) of subscale 2 could not be calculated because of lack of cell filling (Table 6.2). In previous studies using visual estimation methods of the limitation of cervical passive ROM substantial levels of interobserver reliability (K ranged from 0.05 to 0.85 in principal directions and K = 0.54 of all movements together) were reported [21, 38].

The interobserver reliability of the PDS-subscale 3 was somewhat lower compared to others [51].

Compared to all patients in Dutch primary care settings, our patients had less physical impairment (mean PDS 23% versus 58% to 68%) and reported less pain (mean NRS_{pain} 4.0 versus 5.8 to 6.8) but showed a higher neck disability score (mean neck disability index 43% versus 29%) [22, 24, 26, 39, 54]. It is likely that in our patients physical impairment may have played a different role in the biopsychosocial model of disability.

Modified cervical 'non organic signs'

We could not compare the prevalence of positive 'non organic signs' with other neck pain studies but it was similar in patients with chronic low back pain where prevalence of positive 'non organic signs' ranged from 14 to 49% [1, 56]. Likewise it was not possible to compare our kappas for 'non organic signs' with the kappas of the developers of cNOS because prevalence of positive signs was not reported [45]. As a result of prevalence-dependence of the kappas it

is possible that low K values will be found when prevalence of a positive sign is very high or very low. The mean total score of mcNOS was 23%, which is lower than the 35% reported in patients with chronic whiplash associated disorder [52]. This difference may be attributed to differences in operationalization of (m)cNOS, neck pain related disability and pain intensity in the aforementioned studies.

Construct validity of PDS and mcNOS

Although the construct PDS is not fully equal with passive/active ROM the correlation between PDS and NPAD was in line with previous studies which reported a little to fair inverse relationship between passive/active ROM and self-reported neck disability [9, 20, 42, 62]. The correlation between PDS and NRS_{pain} was also in line with other studies in which correlations between passive and active cervical ROM and neck pain ranged from absent to fair [9, 18, 20, 23, 62]. The strength of the association of mcNOS with disability and pain in this study was similar to the association of cNOS with disability and pain in patients with chronic whiplash-associated disorder [52] and similar to the association of NOS with disability and pain in patients with chronic low back pain [35, 56]. Moreover, behavioral signs were stronger related with self-reported disability than physical impairment in this study. Therefore, assessment of PDS and mcNOS may offer relevant information in the biopsychosocial diagnosis of patients with CNP.

Study limitations

There are some limitations of our research. First, the test-retest interval was relatively long and a change in clinical status was not checked upon. In the absence of a gold standard it is unclear which test should be used in determining change in clinical status (actual physical impairment, behavioral signs and self reported disability and pain) during the retest interval. All our patients with CNP were reexamined after the waiting period and in patients with a similar retest interval no substantial change in neck pain related disability was observed [25]. Although the constructs physical impairment and behavioral signs are different from self-reported disability, we assume that the influence of this interval is not substantial. Second, the sample size in this study was adequate for its purpose, although a larger sample is preferred for a validity study.

Other considerations

A strength of this study is that a simple instrument (the PDS) was modified and operationalized, which makes it suitable for use in daily practice for the estimation of the physical impairment. Because the PDS score was substantially lower than was reported in a primary care setting, we expect that the use of the PDS will not be so beneficial in monitoring overall change in tertiary care as in primary care [22]. Furthermore, a mcNOS was constructed and for the first time the prevalence of signs and the interobserver reliability of mcNOS was reported. Based

on what was reported in patients with back pain we expect that in a primary care setting the total score of mcNOS will be substantially lower [55]. As was reported for NOS in patients with chronic low back pain we expect higher intraobserver than interobserver reliability for mcNOS [1]. Future studies are warranted to further develop and investigate the inter- and intraobserver reliability of PDS and mcNOS. It would be interesting to investigate whether a more intensive training would result in a higher interobserver reliability on subscale 2 of the PDS. To optimize the mcNOS the amount of pressure with the deep tenderness and axial loading test should be standardized. Additionally a more precise operationalization of the regional disturbances and a better defining of overreaction tests are desirable. To assess the validity of PDS and mcNOS studies with larger sample sizes are recommendable. In order to test the usefulness of mcNOS as screening tool for patients who require more detailed psychosocial assessment, an in-depth analysis of relationships between mcNOS and psychological and social factors would be needed.

Conclusion

Interobserver reliability of both PDS and mcNOS was acceptable. Clinicians should interpret outcomes of PDS and mcNOS cautiously because outcomes might be biased by examiner differences but also by instability in patient's physical impairment, perceived pain and disability. Moreover, relatively low kappa and absolute agreement values of cervical 'non organic signs' indicate that only the total scores of mcNOS may be of importance when measuring behavioral signs. Behavioral signs were stronger related with self-reported disability than with physical impairment.

KEY POINTS

Findings: An acceptable interobserver reliability of PDS and mcNOS was found in patients with chronic neck pain by experienced clinicians. Relatively low kappa and agreement values of cervical 'non organic signs' indicate that only the total scores of mcNOS may be of importance when measuring behavioral signs.

Implication: mcNOS should be considered as a clinical screening tool that is stronger related with self-reported disability than with PDS.

Caution: The findings are based on the outcomes of measurements in a tertiary rehabilitation setting with younger patients reporting more severe disability than patients in a primary care setting.

REFERENCES

1. Apeldoorn AT, Bosselaar H, Blom-Luberti T, Twisk JW, Lankhorst GJ (2008) The reliability of nonorganic sign-testing and the Waddell score in patients with chronic low back pain. Spine (Phila Pa 1976) 33(7):821–826
2. Apeldoorn AT, Ostelo RW, Fritz JM, van der Ploeg T, van Tulder MW, de Vet HC (2012) The cross-sectional construct validity of the Waddell score. Clin J Pain 28(4):309–317
3. Bergman GJ, Winters JC, Groenier KH, Meyboom-de Jong B, Postema K, van der Heijden GJ (2010) Manipulative therapy in addition to usual care for patients with shoulder complaints: Results of physical examination outcomes in a randomized controlled trial. J Manipulative Physiol Ther 33(2):96–101
4. Bogduk N (1999) The neck. Baillieres Clin Rheumatol 13(2):261–285
5. Bovim G, Schrader H, Sand T (1994) Neck pain in the general population. Spine (Phila Pa 1976) 19(12):1307–1309
6. Bremerich FH, Grob D, Dvorak J, Mannion AF (2008) The neck pain and disability scale: Cross-cultural adaptation into German and evaluation of its psychometric properties in chronic neck pain and C1-2 fusion patients. Spine 33(9):1018–1027
7. Brouwer S, Dijkstra PU, Stewart RE, Goeken LN, Groothoff JW, Geertzen JH (2005) Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. Disabil Rehabil 27(17):999–1005
8. Chen J, Solinger AB, Poncet JF, Lantz CA (1999) Meta-analysis of normative cervical motion. Spine (Phila Pa 1976) 24(15):1571–1578
9. Chiu TT, Lam TH, Hedley AJ (2005) Correlation among physical impairments, pain, disability, and patient satisfaction in patients with chronic neck pain. Arch Phys Med Rehabil 86(3):534–540
10. Cleland JA, Childs JD, Fritz JM, Whitman JM (2006) Interrater reliability of the history and physical examination in patients with mechanical neck pain. Arch Phys Med Rehabil 87(10):1388–1395
11. Cleland JA, Childs JD, Whitman JM (2008) Psychometric properties of the neck disability index and numeric pain rating scale in patients with mechanical neck pain. Arch Phys Med Rehabil 89(1):69–74
12. de Koning CH, van den Heuvel SP, Staal JB, Smits-Engelsman BC, Hendriks EJ (2008) Clinimetric evaluation of active range of motion measures in patients with non-specific neck pain: A systematic review. Eur Spine J 17(7):905–921
13. de Vet HC, Beurskens AJ (1998) Roaming through methodology. VII. Reproducibility of measurements. Ned Tijdschr Geneeskd 142(37):2040–2043
14. Dvorak J, Antinnes JA, Panjabi M, Loustalot D, Bonomo M (1992) Age and gender related normal motion of the cervical spine. Spine (Phila Pa 1976) 17(10 Suppl):S393–398
15. Eriksen HR, Ursin H (2004) Subjective health complaints, sensitization, and sustained cognitive activation (stress). J Psychosom Res 56(4):445-448
16. Fishbain DA, Cole B, Cutler RB, Lewis J, Rosomoff HL, Rosomoff RS (2003) A structured evidence-based review on the meaning of nonorganic physical signs: Waddell signs. Pain Med 4(2):141–181

17. Goolkasian P, Wheeler AH, Gretz SS (2002) The neck pain and disability scale: Test-retest reliability and construct validity. Clin J Pain 18(4):245–250

18. Hagen KB, Harms-Ringdahl K, Enger NO, Hedenstad R, Morten H (1997) Relationship between subjective neck disorders and cervical spine mobility and motion-related pain in male machine operators. Spine (Phila Pa 1976) 22(13):1501–1507

19. Hee HT, Whitecloud TS,3rd, Myers L, Roesch W, Ricciardi JE (2002) Do worker's compensation patients with neck pain have lower SF-36 scores? Eur Spine J 11(4):375–381

20. Hermann KM, Reese CS (2001) Relationships among selected measures of impairment, functional limitation, and disability in patients with cervical spine disorders. Phys Ther 81(3):903–914

21. Hoppenbrouwers M, Eckhardt MM, Verkerk K, Verhagen A (2006) Reproducibility of the measurement of active and passive cervical range of motion. J Manipulative Physiol Ther 29(5):363–367

22. Hoving JL, de Vet HC, Koes BW, Mameren H, Deville WL, van der Windt DA, Assendelft WJ, Pool JJ, Scholten RJ, Korthals-de Bos IB, Bouter LM (2006) Manual therapy, physical therapy, or continued care by the general practitioner for patients with neck pain: Long-term results from a pragmatic randomized clinical trial. Clin J Pain 22(4):370–377

23. Jordan A, Manniche C, Mosdal C, Hindsberger C (1998) The Copenhagen neck functional disability scale: A study of reliability and validity. J Manipulative Physiol Ther 21(8):520–527

24. Jorritsma W, de Vries GE, Dijkstra PU, Geertzen JH, Reneman MF (2012) Neck pain and disability scale and neck disability index: Validity of Dutch language versions. Eur Spine J 21(1):93–100

25. Jorritsma W, de Vries GE, Geertzen JH, Dijkstra PU, Reneman MF (2010) Neck pain and disability scale and the neck disability index: Reproducibility of the Dutch language versions. Eur Spine J 19(10):1695–1701

26. Koes BW, Bouter LM, van Mameren H, Essers AH, Verstegen GM, Hofhuizen DM, Houben JP, Knipschild PG (1992) A blinded randomized clinical trial of manual therapy and physiotherapy for chronic back and neck complaints: Physical outcome measures. J Manipulative Physiol Ther 15(1):16–23

27. Kuhlman KA (1993) Cervical range of motion in the elderly. Arch Phys Med Rehabil 74(10):1071–1079

28. Landers MR, Creger RV, Baker CV, Stutelberg KS (2008) The use of fear-avoidance beliefs and nonorganic signs in predicting prolonged disability in patients with neck pain. Man Ther 13(3):239–248

29. Main CJ, Waddell G (1998) Behavioral responses to examination. A reappraisal of the interpretation of "nonorganic signs". Spine (Phila Pa 1976) 23(21):2367–2371

30. Makela M, Heliovaara M, Sievers K, Impivaara O, Knekt P, Aromaa A (1991) Prevalence, determinants, and consequences of chronic neck pain in Finland. Am J Epidemiol 134(11):1356–1367

31. Malmstrom EM, Karlberg M, Fransson PA, Melander A, Magnusson M (2006) Primary and coupled cervical movements: The effect of age, gender, and body mass index. A 3-dimensional movement analysis of a population without symptoms of neck disorders. Spine (Phila Pa 1976) 31(2):E44–50

32. Nachemson A, Vingard E (2000) Assessment of patients with neck and back pain: A best-evidence synthesis. In: Nachemson AL, Jonsson E, eds. Neck and Back Pain. Philadelphia: Lippincott Williams & Wilkens 189–236

33. Nordin M, Carragee EJ, Hogg-Johnson S, Weiner SS, Hurwitz EL, Peloso PM, Guzman J, van der Velde G, Carroll LJ, Holm LW, Cote P, Cassidy JD, Haldeman S (2009) Assessment of neck pain and its associated disorders: Results of the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. J Manipulative Physiol Ther 32(2 Suppl):S117–140

34. Norlander S, Gustavsson BA, Lindell J, Nordgren B (1997) Reduced mobility in the cervico-thoracic motion segment--a risk factor for musculoskeletal neck-shoulder pain: A two-year prospective follow-up study. Scand J Rehabil Med 29(3):167–174

35. Novy DM, Collins HS, Nelson DV, Thomas AG, Wiggins M, Martinez A, Irving GA (1998) Waddell signs: Distributional properties and correlates. Arch Phys Med Rehabil 79(7):820–822

36. Ombreght L, Bisschop P, Ter Veer HJ (1995) A system of orthopaedic medicine. London: WB Saunders Company Ltd.

37. Picavet HS, Schouten JS (2003) Musculoskeletal pain in The Netherlands: Prevalences, consequences and risk groups, the DMC(3)-study. Pain 102(1-2):167–178

38. Pool JJ, Hoving JL, de Vet HC, van Mameren H, Bouter LM (2004) The interexaminer reproducibility of physical examination of the cervical spine. J Manipulative Physiol Ther 27(2):84–90

39. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC (2007) Minimal clinically important change of the neck disability index and the numerical rating scale for patients with neck pain. Spine (Phila Pa 1976) 32(26):3047–3051

40. Portney LG, Watkins MP (2000) Foundations of clinical research 2nd edition prentice hall health. Upper Saddle River, New Jersey

41. Reesink DD, Jorritsma W, Reneman MF (2007) Basis for a functional capacity evaluation methodology for patients with work-related neck disorders. J Occup Rehabil 17(3):436--449

42. Riddle DL, Stratford PW (1998) Use of generic versus region-specific functional status measures on patients with cervical spine disorders. Phys Ther 78(9):951–963

43. Schellingerhout JM, Heymans MW, Verhagen AP, Lewis M, de Vet HC, Koes BW (2010) Prognosis of patients with nonspecific neck pain: Development and external validation of a prediction rule for persistence of complaints. Spine (Phila Pa 1976) 35(17):E827–835

44. Sjors A, Larsson B, Persson AL, Gerdle B (2011) An increased response to experimental muscle pain is related to psychological status in women with chronic non-traumatic neck-shoulder pain. BMC Musculoskelet Disord 12:230

45. Sobel JB, Sollenberger P, Robinson R, Polatin PB, Gatchel RJ (2000) Cervical nonorganic signs: A new clinical tool to assess abnormal illness behavior in neck pain patients: A pilot study. Arch Phys Med Rehabil 81(2):170–175

46. Sobel JS, Kremer I, Winters JC, Arendzen JH, de Jong BM (1996) The influence of the mobility in the cervicothoracic spine and the upper ribs (shoulder girdle) on the mobility of the scapulohumeral joint. J Manipulative Physiol Ther 19(7):469–474

47. Sterling M (2008) Testing for sensory hypersensitivity or central hyperexcitability associated with cervical spine pain. J Manipulative Physiol Ther 31(7):534–539

48. Streiner D, Norman G (1995) Health measurement scales: A practical guide to their development and use. 2nd ed. Oxford: Oxford University Press

49. Stucki G (2005) International classification of functioning, disability, and health (ICF): A promising framework and classification for rehabilitation medicine. Am J Phys Med Rehabil 84(10):733–740

50. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60(1):34–42

51. Terwee CB, de Winter AF, Scholten RJ, Jans MP, Devilllé W, van Schaardenburg D, Bouter LM (2005) Interobserver reproducibility of the visual estimation of range of motion of the shoulder. Arch Phys Med Rehabil 86(7):1356–1361

52. Vernon H, Guerriero R, Soave D, Kavanaugh S, Puhl A, Reinhart C (2011) The relationship between self-rated disability, fear-avoidance beliefs, and nonorganic signs in patients with chronic whiplash-associated disorder. J Manipulative Physiol Ther 34(8):506–513

53. Vernon H, Proctor D, Bakalovski D, Moreton J (2010) Simulation tests for cervical nonorganic signs: A study of face validity. J Manipulative Physiol Ther 33(1):20–28

54. Vos CJ, Verhagen AP, Koes BW (2006) Reliability and responsiveness of the Dutch version of the neck disability index in patients with acute neck pain in general practice. Eur Spine J 15(11):1729–1736

55. Waddell G, McCulloch JA, Kummel E, Venner RM (1980) Nonorganic physical signs in low-back pain. Spine (Phila Pa 1976) 5(2):117–125

56. Waddell G, Morris EW, Di Paola MP, Bircher M, Finlayson D (1986) A concept of illness tested as an improved basis for surgical decisions in low-back disorders. Spine (Phila Pa 1976) 11(7):71271–71279

57. Wheeler AH, Goolkasian P, Baird AC, Darden BV (1999) Development of the neck pain and disability scale. Item analysis, face, and criterion-related validity. Spine 24(13):1290–1294

58. WHO (2001) ICF: International classification of functioning, disability, and health. Geneva: World Health Organization

59. Williams MA, McCarthy CJ, Chorti A, Cooke MW, Gates S (2010) A systematic review of reliability and validity studies of methods for measuring active and passive cervical range of motion. J Manipulative Physiol Ther 33(2):138–155

60. Williamson A, Hoggart B (2005) Pain: A review of three commonly used pain rating scales. J Clin Nurs 14(7):798–804

61. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel M (2002) French translation and validation of 3 functional disability scales for neck pain. Arch Phys Med Rehabil 83(3):376–382

62. Ylinen J, Takala EP, Kautiainen H, Nykänen M, Häkkinen A, Pohjolainen T, Karppi SL, Airaksinen O (2004) Association of neck pain, disability and neck pain during maximal effort with neck muscle strength and range of movement in women with chronic non-specific neck pain. Eur J Pain 8(5):473–478

APPENDIX 6.1 PHYSICAL DYSFUNCTION SEVERITY

Name patient: _____

Date of birth: _____

Date of examination: _____

Position: Patient is sitting actively on a chair or examination table

			SCORE
Pain during active cervical movement		**Subscale 1 (ranges 0–6)**	
	0	1	
Flexion	no	yes	
Extension	no	yes	
Lateral flexion (left)	no	yes	
Lateral flexion (right)	no	yes	
Rotation (left)	no	yes	
Rotation (right)	no	yes	

Limitation of passive cervical movement			**Subscale 2 (ranges 0–18)**		
	0 (no)	1 (light)	2 (moderate)	3 (severe)	
Flexion	60–70°	45–60°	30–45°	0–30°	
Extension	60–70°	45–60°	30–45°	0–30°	
Lateral flexion (left)	40–45°	30–40°	20–30°	0–20°	
Lateral flexion (right)	40–45°	30–40°	20–30°	0–20°	
Rotation (left)	80–90°	60–80°	40–60°	0–40°	
Rotation (right)	80–90°	60–80°	40–60°	0–40°	

Limitation of passive shoulder flexion			**Subscale 3 (ranges 0–6)**		
	0 (no)	1 (light)	2 (moderate)	3 (severe)	
Left	170–180°	128–170°	85–128°	0–85°	
Right	170–180°	128–170°	85–128°	0–85°	

TOTAL SCORE (ranges 0–30)	

APPENDIX 6.2 OPERATIONALIZATIONS OF MODIFIED CERVICAL NONORGANIC SIGNS

Sign	Test	Criteria for a positive test
I Tenderness		
1. Superficial	Palpation of skinfolds at both sides of the cervical and upper thoracic spine region	Skinfolds are tender to light pinching over a widespread area
2. Deep	Posterior to anterior pressure centrally over the spinous processus of the vertebrae of the cervical and upper thoracic spine (C2–T5)	Bony tenderness over a widespread area
II Simulation		
3. Axial loading	Pressure of a few pounds on the shoulders	Axial loading produces (more) neck pain
4. Simulated rotation	Seated passive trunk rotation with the hands of the patient folded behind the neck	Trunk rotation produces (more) neck pain
III Distraction		
5. Extension	Looking upward when lying prone sustained on flexed elbows	Marked difference (approximately 20°) between formal active extension and extension on distraction
6/7. Rotation	Lying prone with the head rotated and lifting the extended ipsilateral arm out of 90° abduction in the shoulder	Marked difference (approximately 30°) between formal active rotation and rotation on distraction
IV Regional disturbances		
8. Motor loss	Formal manual muscle testing, of shoulder and elbow muscle groups	Widespread weakness that does not fit in a neuro-anatomic pattern; the key-finding is jerky/"giving way" weakness
9. Sensory loss	Light touch or pinprick	Patient reports diminished sensation in a pattern that does not correspond to a specific dermatome of a nerve root(s) or peripheral nerve(s)
V Overreaction		
10. Disproportional verbalization/sighing	Observation	Examiner interprets the reaction of the patient as "overreacting", during the examination
11. Grimacing due to pain		
12. Rubbing/clutching the affected area		
13. Stiff, rigid or slow movement		

Chapter 7

Basis for a functional capacity evaluation methodology for patients with work-related neck disorders

David D. Reesink
Wim Jorritsma
Michiel F. Reneman

J Occup Rehabil (2007) 17:436–449

ABSTRACT

Background: Neck pain is a common musculoskeletal complaint and a relationship with reduced work-related functional capacity is assumed. A validated instrument to test functional capacity of patients with neck pain is unavailable. The objective of this study was to develop a Functional Capacity Evaluation (FCE), which is content valid for determining functional capacity in patients with work related neck disorders (WRND).

Methods: A review of epidemiological review literature was conducted to identify physical risk factors for WRND.

Results: Evidence was found that physical risk factors contribute in development of WRND. Physical risk factors were related to repetitive movements, forceful movements, awkward positions and static contractions of the neck or the neck/shoulder region. An FCE was designed based on the risk factors identified. Eight tests were selected to cover all risk factors: repetitive side reaching, repetitive reaching overhead, static overhead work, front carry, forward static bend neck, overhead lift and the neck strength test. Content validity of this FCE was established by providing the rationale, specific objectives and operational definitions of the FCE.

Conclusions: Further research is needed to establish reliability and other aspects of validity of the neck-FCE.

INTRODUCTION

Neck pain is a common musculoskeletal complaint [21]. Among the general population in the Netherlands, the 12-month prevalence of neck pain was 31.4% [55]. Among the general working population in the Netherlands the 12-month prevalence figures for low back, neck and shoulder complaints, were respectively 44.5, 28.5, and 27.3% [32]. Prevalence in the general population in other countries are similar [5, 16, 20, 46, 54]. The 1-year prevalence of neck pain varies among different occupational groups from 6 to 76%, with higher values for female workers [9, 16, 39, 61, 64, 65]. The pathological basis for neck pain is unclear in approximately 80% of the cases [14]; the term "non-specific" is applied to these cases [47]. Neck pain is often concurrent with shoulder pain [55, 67]. Dysfunction of the cervico-thoracic region strongly predicts development of shoulder complaints [50, 51]. A limited shoulder/arm function, can affect the prevalence of neck pain. Neck pain can be a substantial burden on society, because it is related to work disability, unemployment and insurance claims. The majority of these costs are not related to health care, but are due to sick leave, disability and loss of productivity [15].

Work related neck pain (WRNP) is a multi-factorial disorder among works. It may in part be caused, aggravated, accelerated or exacerbated by occupational exposures, and may be related to impaired work capacity [71]. The definition of WRNP is based on a definition of work related disability by the World Health Organization [1]. In the WHO domain of capacity, the definition of WRNP is characterized by a diminishing work capacity. Work capacity represents the limits of the anatomical, physiological and psychological systems in work-related parameters, which include job demands specific to a job [2]. The work capacity is performance based determined through a functional capacity evaluation (FCE). An FCE aims to measure an individual's physical capability to perform work related activities [35, 36, 45]. FCEs are used in work rehabilitation programs, for disability determination, return to work recommendations and in medico-legal issues.

Most research on functional capacity or disability regarding patients with WRND is conducted through questionnaires based on self-report. Comparative research between a patient's self-report, clinical examination (physician) and FCE has shown little similarity and low correlation in patients with chronic low back pain [18]. Self-reported limitations were considerably higher than those derived from clinical examination or from FCE. Limitations based on clinical examinations were higher than those derived from FCE. Each method deals with a different perspective; self-report instruments may reveal a patient's perception of his own performance, clinical examination reveals a physician's perception of a patient's performance and functional testing (FCE) reveals the actual performance of a patient in a standardized assessment setting. Balogh et al. [10] and Hansson et al. [28] have shown that subjects with neck/shoulder complaints rated their work exposure higher than those without, although they in fact showed lower direct measured exposure. Thus an FCE specific for patients with neck pain is needed. At this moment, however, no validated performance based instrument has been described in literature.

The objective of this study was to develop an FCE, which is content valid for determining functional capacity in patients with WRND. Content validity is the degree to which test items represent the performance domain the test is intended to measure [33]. To determine content validity it is necessary to establish rational grounds for the test, to provide operational definitions of the test variables and to identify the specific objectives of the test [56]. As previously described in the Journal of Occupational Rehabilitation [59] a review of epidemiological literature was conducted to identify the rational grounds for physical risk factors related to work related upper limb disorders. There is a widely supported idea that there are certain types of work are linked to WRND. Consequently, different work related risk factors should be described. The identified physical risk factors form the basis for a WRND specific FCE. Functional tests will be selected to cover these risk factors.

Besides physical risk factors, also non-physical factors are known to influence WRND and sickness absence [22]. Psychological and personality traits, health beliefs, environmental and social circumstances at work or at home, coping resources, mood, and psychopathology are potentially important in the development or sustenance of WRND [22]. However, this study focuses on physical aspects only, with the objective that this neck-FCE will be content valid in measuring work-related functional capacity of patients with WRND.

METHODS

Literature review

The electronic databases MEDLINE (1966–October 2006), EMBASE (1966–October 2006) and Cochrane Library (1966–October 2006) were systematically searched to identify work related physical risk factors in epidemiologic review studies. These databases were searched with the following keywords: work related neck disorders, occupational cervicobrachial disorders (and OCD), work related upper extremity disorders (and WRUED), upper extremity musculoskeletal disorders (and UEMS), work-related upper limb disorders (and WRULD), upper limb disorders (and ULD), repetitive strain injury (and RSI), cumulative trauma disorders (and CTD), musculoskeletal disorders (and MSD), and work related cumulative trauma disorders (and WRCTD). These searches were performed in combination with the terms neck or neck/shoulder. Reference lists of the selected studies were screened for additional relevant studies. Relevant review-studies with references to studies already reviewed were excluded to avoid overlap. Full text review articles were included if they identified a work-related risk factor for neck and neck/shoulder pain. Reviews that dealt with neck/shoulder pain were only included if the shoulder was related to neck pain. Only reviews written in English were included.

Risk factors operational definition and selection

Work related physical risk factors were selected from literature reviews. The content of similar risk factors' were checked for correspondence. To ensure that each similar risk factor definition had the same content, they were operationally defined. The risk factors were operationally defined, with regards to: (1) nature of the risk factor; (2) point within the body being considered [26, 72]; (3) total duration, intensity [26, 72], or frequency of the exposure; and (4) the axis (longitudinal, sagittal, transversal) where the movements or posture of the neck or shoulder was performed. The type of movement or posture defined the nature of a risk factor. Correspondence between risk factors was established if the first three operational definitions showed conformity. When no consensus was reached between the risk factors, they were described separately. A risk factor was excluded when it dealt solely with duration, intensity, or frequency in relation to a specific job. When risk factors were undefined in the reviews, they were suppleted with definitions of other studies dealing with WRND risk factors. Risk factors were included in the FCE when three non-systematic and non-critical reviews, or two systematic and critical reviews, or two systematic reviews concluded that there was a relationship between neck pain and that risk factor.

Test design and selection

An FCE was designed which covered the included physical risk factors as they were operationally defined. Tests were selected based on four criteria:

(1) Whether the test measured at a functional level. When risk factors could not be covered by tests at a functional level, tests measuring at a non-functional level were selected.
(2) Tests were selected when they covered one risk factor. When not available, tests were selected to cover two or more risk factors simultaneously.
(3) If the risk factor was defined on multiple axes, tests were selected to cover these. When not available, tests measuring one axis were selected.
(4) Guidelines in hierarchical order for functional capacity evaluation presented by Hart et al. [29];
    (a) Safe: Given the known characteristics of the evaluee, the procedure should not be expected to lead to injury,
    (b) Reliability: the test score should be dependable across evaluators, evacuee's, and the date or time of test administration,
    (c) Validity: The interpretation of the test score should be able to predict or reflect the evacuee's performance in target work setting,
    (d) Practicality: the cost of the test procedure should be reasonable and customary,
    (e) Utility: the usefulness of the procedure is the degree to which it meets the needs of the evaluee, referrer, and payer.

Based on these criteria, existing tests as documented in the Workwell FCE protocol [34] were selected. If unavailable, existing tests documented in other literature were selected. Tests were modified from existing tests or test were self-designed, when risk factors could not be covered as operational defined.

RESULTS

Literature review

Nine reviews met the inclusion criteria, four of which were non-systematic and noncritical reviews, two were systematic reviews and four were systematic and critical reviews. A well performed critical review published in a handbook was also included [49]. The results of the literature review are summarized in Table 7.1.

Risk factors operational definition and selection

Risk factor operational definition

There is uniformity between the reviews in the nature of the risk factors and the point of the body being considered. No review reported values of frequency, intensity or total duration. There is consensus about the content of the same risk factors in the three operational definitions (nature, body part considered and exposure). The reviews are uniform in describing the nature of the risk factors, and which body part is being considered. None of the reviews described detailed content of the operational definitions of the risk factors. The nature of the risk factors are operationally defined in the next paragraph.

Table 7.1 Summary of the reviews categorized according to work related physical risk factors in the neck/shoulder region [6, 11, 19, 24, 34, 48, 49, 57, 73]

Risk factor	References									
	[34]	[49]	[58a]	[58b]	[11]	[48]	[73]	[6]	[19]	[24]
Repetitive movements	C	–	C	C	S	R	C	R	–	–
Awkward positions	C	S	C	C	–	–	C	R	R	R
Forceful movements	–	–	–	C	S	–	C	R	–	–
Static contractions	C	–	C	C	S	R	–	–	–	R

S, systematic review; C, systematic and critical review; R, non-systematic and non-critical review; –, no evidence described; *[a] review dedicated to neck; *[b] review dedicated to neck/shoulder.

Nature of the risk factors

Repetitive movements: repeated or cyclical neck movements, or repeated arm or shoulder motions that generate load to the neck/shoulder region e.g., trapezium muscle [11]. A cycle is a short-term trend that is expected to reverse.

Awkward position: a combination of forceful and repetitive movements in an extreme position of the neck/shoulder region.

Forceful movements: loads to the neck and neck/shoulder, or described exposure as strenuous work involving the upper extremity that generates load to the neck/shoulder muscles [40].

Static contractions: long-term exposure or static posture that generates load on the neck/shoulder muscles or other prolonged isometric contractions of the neck/shoulder muscles.

Risk factor selection

On the basis of the reviews and the operational process, the following risk factors were included using the inclusion criteria for neck/shoulder region: repetitive movements, awkward position, forceful movements, and static contractions. For each risk factor, operational values are presented in Table 7.2.

Test design

An FCE was designed to cover the risk factors for the neck or neck/shoulder region. Tests selected are presented in Table 7.3. An FCE consisting of seven tests was designed to cover the risk factors. Four functional tests were included from existing tests from the Workwell FCE protocol [34]: repetitive side reaching, static overhead work, front carry and overhead lift. One existing functional test from this FCE was modified: repetitive reaching overhead. One direct performance test described in literature was included: the neck strength test. One functional test was self-designed: forward static neck bend test sitting.

The work related risk factors were covered by the tests in Table 7.3. The content of these tests is described below. The operational definitions of Table 7.2 have been used to specify the content of these tests.

Repetitive side reaching test

Objective: Fast repetitive movements of the upper extremity. Materials: 30 small objects in bowls, shelf or table of maximal wing span apart, positioned at midthoracic height. Procedure: Sitting with bowls apart at maximal wing span. Remove marbles horizontally at table height from left to right with left arm as fast as possible (and vice versa). Time needed to remove 30 marbles is scored (s). Test–retest reliability: in patients with chronic low back pain (CLBP), Intraclass Correlation (ICC) = 0.45–0.64 [17]; in healthy subjects ICC = 0.74–0.76 [66] and ICC = 0.54–0.72

Table 7.2 Risk factors operationally defined in: nature of the risk factor (summarized), total duration, intensity, or frequency of the exposure, and the axis (i.e. longitudinal, transversal, sagittal) were the movement or posture of the neck or shoulder was performed [4, 7, 8, 38, 40-42, 44, 52, 53, 62, 63, 69]

Nature of risk factor	Physical exposure (A)			Neck axis (B)			Shoulder axis (B)		
	Intensity (%MVC)	Duration (time)	Frequency (cycles per min)	Longitudinal (degrees)	Transversal (degrees)	Sagittal (degrees)	Longitudinal (degrees)	Transversal (degrees)	Sagittal (degrees)
Repetitive movements	<10 [40, 52]	–	>15 [38, 52]	X [4, 41, 42]	–	X [4, 41, 42]	–	X	X
Awkward positions	–	X	–	>45 [44]	–	>20 [8, 38]	–	>60 [38]	>90 [7, 63, 69]
Forceful movements	>10 [4, 41, 42, 52]	–	–	–	X	X	–	X	X
Static contractions	X	X[53, 62]	–	>45 [44]	–	>20 [8, 38]	–	>60 [38]	>90 [7, 63, 69]

X, item consistently referred to, but inconsistently operational defined; MVC, Maximal voluntary contraction; <, less than; >, more than; –, no operational value defined in literature; (A) frequency, intensity, and total duration of the exposure; and (B) the plane or axis where the movements or posture is performed.

Table 7.3 Content summary of the FCE based on included risk factors for neck/shoulder region

Tests:	Repetitive side reaching	Repetitive reaching overhead	Static overhead work	Front carry	Forward static bend neck	Overhead lift	Neck strength test
Risk factors							
Repetitive movements	√	√	–	–	–	–	–
Awkward positions	√	–	√	–	√	–	–
Forceful movements	–	–	–	√	–	√	√
Static contractions	–	–	√	√	√	–	–

√-marked cell, coverage of the risk factor by different tests.

[58]. The risk factors covered by this test are repetition and awkward positions in the neck and shoulder both in the longitudinal axis.

Repetitive overhead reaching test

Objective: Fast repetitive movements of the upper extremity. Materials: 20 marbles and two bowls with a 14-cm diameter positioned at crown height and at table height. Procedure: Test adjusted from the IWS FCE dynamic bending. Standing in front of the bowls and moving the marbles as fast as possible from table height to crown height. Test–retest reliability: Test not described in literature. The risk factor covered by this test is repetition in the neck and shoulder in the sagittal axis.

Static overhead work test

Objective: Static holding time of shoulder and neck musculature. Materials: Aluminum plate adjustable in height with 20 holes, bolts and nuts and two cuff weights of 1.0 kg each. Procedure: Standing with hands at crown height, manipulating nuts and bolts wearing cuff weights around the wrists. The time that position is held will be measured in seconds. Test–retest reliability: in healthy subjects ICC = 0.90 [66]. The risk factors covered by this test are awkward position of the shoulder on the transversal and sagittal axis and static contractions in the neck and shoulder both on the transversal and sagittal axis.

Front carry test

Objective: Carry weight receptacle in two-handed manner at waist level. Materials: Plastic receptacle (40 x 30 x 26 cm). Weights of 1.0, 2.0, and 4.0 kg. Procedure: Begin with suitable weight, carry 20 m up and back within 90 s. Increase weights in 4–5 steps until maximum is reached.

Test–retest reliability: In CLBP ICC = 0.81 [31] and 0.87–90 [25]; and in healthy adults ICC = 0.84 [58]. The risk factors covered by this test are forceful movements and static contractions in the neck and shoulder both around the sagittal, transversal and longitudinal axes.

Forward static neck bend test sitting

Objective: Sitting with 20–45 degrees neck flexion and manipulation nut/bolts on table height. Materials: Nut/bolts loaded helmed (3 kg). Procedure: Sitting with 20–45 degrees neck flexion with loaded helmed manipulation nuts/bolts on table height. The time that position is held will be measured in seconds. Test–retest reliability: Test not described in literature. The risk factors covered by this test are awkward positions and static contractions in the neck around the sagittal axis.

Overhead lift test

Objective: Functional strength of shoulder and arm musculature. Materials: Plastic receptacle (40 x 30 x 26 cm). A wall mounted system with adjustable shelves and weights of 1.0, 2.0, and 4.0 kg. Procedure: Five lifts from waist to crown height and vice versa within 90 s in standing position. Increase weight in 4–5 steps until maximum is reached. Test–retest reliability: In CLBP ICC = 0.87 [60] and 0.81–84 [25]; in healthy subjects ICC = 0.89 [58] and 0.92 [66]. The risk factor covered by this test is forceful movements in the shoulder on the sagittal axis.

Neck strength test

Objective: Muscular capacity in isometric neck flexion or extension and lateral flexion left and right. Materials: A hand-held dynamometer (type MicroFET2®) with a stabilization device and a swivel chair. Procedure: Subject seated on the chair with hands on the hips and feet supported on the floor. The position of the dynamometer is then adjusted to be orthogonal directed against the forehead (in flexion), occipital (in extension) and parietal (lateral flexion) regions. The output was measured by mean of the best 2 of 5 maximal isometric contractions. Test–retest reliability: In healthy subjects the reliability is good [3, 12, 13, 23, 30]. The risk factor that is covered by this test is forceful movements in the neck on the sagittal and transversal axes.

DISCUSSION

The objective of this study was to develop an FCE, which is content valid for determining functional capacity in patients with WRND. The importance of this study lies in the fact that previously no such neck-FCE has been described in literature. This neck-FCE is based on the present understanding of physical work related risk factors for the neck. Physical work related factors were identified from epidemiologic review literature for the specific purpose of development of a neck-FCE. The literature search resulted in the selection of nine review studies. These WRND studies were the rational basis for a content valid FCE and were used to identify the specific objectives of the

instrument [56]. Studies might have been missed because the reviews on risk factors reported on several outcome measures, of which neck pain was one. Keywords indicating the neck were missing and as a result studies may have been missed. This was corrected by the overlap of the reviewed studies in the reviews. This means that the same studies were included in different reviews. A substantial overlap between review studies made it necessary to exclude three reviews [26, 27, 70].

A disadvantage in the selection of risk factors specifically by use of reviews is that relevant information is lost because of generalization. Checks were necessary to establish whether or not the generalizations in risk factors between the different reviews were defined in the same manner. This was achieved by operationally defining the risk factors into definitions that could be compared. Consensus was reached if the first three operationalised definitions (nature, body part considered and exposure of the risk factor) corresponded. If the operational definitions were not described in the review studies, they were suppleted from other studies. These supplements were drawn from studies that were reviewed by the selected reviews and other WRND studies. This selection was neither systematic nor critical. A weakness in this method is that this supplement is not exhaustive; however it does clarify the boundaries between the risk factors. A strength of this study was that operational risk factors could be easily and objectively modified into test items which covered them. For example, a description of the risk factor 'awkward positions' is not testable until it is operationalised. An important limitation of the literature was the overall absence of definitions of risk factors. There is great uniformity in describing the nature of the risk factors, but only one review described content of the risk factors [6]. It is not certain that reviews dealt with the same content because no definitions were presented. So, there may be overlap or selection bias between the risk factors.

Content validity is that kind of validity which measures the degree to which test items represent the performance domain the test is intended to measure [33]. Content validity is usually determined by a panel of experts or by knowledge of the normal practices that examine the relationship between test objectives and test items [37, 68]. The FCE in this study is based on the knowledge of present understanding regarding work related risk factors for the neck. Literature research was preferred over an expert panel, to determine risk factors for WRND. This was done because research based information was gathered. This can be considered as a strength of this study. The Dictionary of Occupational Titles (DOT) has been used to assess the content validity of existing FCEs by examining how well the evaluation covers the physical demands defined in the DOT [33, 43]. Even though the DOT taxonomy is widely used, it has not been validated. Post-hoc content analysis of our neck-FCE revealed that the physical demands, lifting, reaching and handling, described in the DOT, are also included in this FCE.

However, the question remains if for the individual WRND-patients with a particular occupation all physical risk factors that might have contributed to its development are covered by the WRND FCE. Furthermore, we do not know if the patient's neck pain is more inconvenient when performing under certain work related activities, or that these work related activities are the

primary factors for the development of the WRND. Another important aspect is that patients underestimate their work related functional capacity and therefore possibly have (longer) sick leave. The authors are aware of the fact that psychosocial factors are also of importance to sickness absence. Nevertheless, we assume that this neck-FCE is capable to measure the work-related functional capacity of patients with WRND.

WRNDs are usually defined in multidimensional terms [22]. In spite of the relevance of non-physical risk factors, the objective of this study was to use physical risk factors only. For clinical use, it is recommended that the WRND FCE is combined with medical and psychosocial evaluations to assess these other dimensions. Validation of this WRND FCE has started with this study. Further steps in test development will be, determining test–retest reliability and construct and criterion related validity [33], in particular the determination of (reduced) capacity in return to work situations.

Conclusion

The review has provided evidence that physical risk factors contribute in the development of WRND. These physical risk factors were repetitive movements, forceful movements, awkward positions and static contractions for the neck or neck/shoulder region. The neck-FCE based on this review provides evidence for the content validity.

Acknowledgements

The authors thank M.T. Wolters for critically reviewing prior versions of this manuscript.

REFERENCES

1. Anonymous (1973) Environmental and health monitoring in occupational health. 535:1–48
2. Abdel-Moty E, Compton R, Steele-Rosomoff R, Rosomoff L, Khalil TM (1996) Process analysis of functional capacity assessment. Journal of Back and Musculoskeletal Rehabilitation 6:223–236
3. Agre JC, Magness JL, Hull SZ, Wright KC, Baxter TL, Patterson R, Stradel L (1987) Strength testing with a portable dynamometer: reliability for upper and lower extremities. Arch Phys Med Rehabil 68:454–458
4. Andersen JH, Kaergaard A, Frost P, Thomsen JF, Bonde JP, Fallentin N, Borg V, Mikkelsen S (2002) Physical, psychosocial, and individual risk factors for neck/shoulder pain with pressure tenderness in the muscles among workers performing monotonous, repetitive work. Spine (Phila Pa 1976) 27:660–667
5. Andersson HI, Ejlertsson G, Leden I, Rosenberg C (1993) Chronic pain in a geographically defined general population: studies of differences in age, gender, social class, and pain localization. Clin J Pain 9:174–182
6. Ariens GA, van Mechelen W, Bongers PM, Bouter LM, van der Wal G (2000) Physical risk factors for neck pain. Scand J Work Environ Health 26:7–19

7. Ariens GA, Bongers PM, Douwes M, Miedema MC, Hoogendoorn WE, van der Wal G, Bouter LM, van Mechelen W (2001) Are neck flexion, neck rotation, and sitting at work risk factors for neck pain? Results of a prospective cohort study. Occup Environ Med 58:200–207

8. Ariens GA, Bongers PM, Hoogendoorn WE, van der Wal G, van Mechelen W (2002) High physical and psychosocial load at work and sickness absence due to neck pain. Scand J Work Environ Health 28:222–231

9. Ariens GAM, Borghouts JAJ, Koes BW (1999) The epidemiology of pain. In: Crombie IK (ed) Neck pain. IASP Press, Seattle, pp 235–255

10. Balogh I, Orbaek P, Ohlsson K, Nordander C, Unge J, Winkel J, Hansson GA, Malmo Shoulder/Neck Study Group (2004) Self-assessed and directly measured occupational physical activities--influence of musculoskeletal complaints, age and gender. Appl Ergon 35:49–56

11. Bernard BP (1997) Musculoskeletal Disorders and Workplace Factors. U.S. Department of Health and Human Services, Cincinnati (OH)

12. Bohannon RW (1986) Test-retest reliability of hand-held dynamometry during a single session of strength assessment. Phys Ther 66:206–209

13. Bohannon RW, Andrews AW (1987) Interrater reliability of hand-held dynamometry. Phys Ther 67:931–933

14. Bongers PM, Kremer AM, ter Laak J (2002) Are psychosocial factors, risk factors for symptoms and signs of the shoulder, elbow, or hand/wrist?: A review of the epidemiological literature. Am J Ind Med 41:315–342

15. Borghouts JA, Koes BW, Vondeling H, Bouter LM (1999) Cost-of-illness of neck pain in The Netherlands in 1996. Pain 80:629–636

16. Bovim G, Schrader H, Sand T (1994) Neck pain in the general population. Spine (Phila Pa 1976) 19:1307–1309

17. Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens JM, Goeken LN (2003) Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in patients with chronic low back pain. J Occup Rehabil 13:207–218

18. Brouwer S, Dijkstra PU, Stewart RE, Goeken LN, Groothoff JW, Geertzen JH (2005) Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. Disabil Rehabil 27:999–1005

19. Buckle PW, Devereux JJ (2002) The nature of work-related neck and upper limb musculoskeletal disorders. Appl Ergon 33:207–217

20. Cote P, Cassidy JD, Carroll L (1998) The Saskatchewan Health and Back Pain Survey. The prevalence of neck pain and related disability in Saskatchewan adults. Spine (Phila Pa 1976) 23:1689–1698

21. Ferrari R, Russell AS (2003) Regional musculoskeletal conditions: neck pain. Best Pract Res Clin Rheumatol 17:57–70

22. Feuerstein M, Shaw WS, Nicholas RA, Huang GD (2004) From confounders to suspected risk factors: psychosocial factors and work-related upper extremity disorders. J Electromyogr Kinesiol 14:171–178

23. Fiebert IM, Roach KE, Yang SS, Dierking LD, Hart FE (1999) Cervical range of motion and strength during resting and neutral head postures in healthy young adults. J Back Musculoskel Rehab 12 (3):165–178

24. Gerr F, Marcus M, Monteilh C (2004) Epidemiology of musculoskeletal disorders among computer users: lesson learned from the role of posture and keyboard use. J Electromyogr Kinesiol 14:25–31

25. Gross DP, Battie MC (2002) Reliability of safe maximum lifting determinations of a functional capacity evaluation. Phys Ther 82:364–371

26. Hagberg M, Silverstein BA, Wells RV, Smith MJ, Hendrick HW, Carayon P, Pérusse M (1995) Work related musculoskeletal disorders (WMSDs): A reference book for prevention. Taylor and Francis, London

27. Hales TR, Bernard BP (1996) Epidemiology of work-related musculoskeletal disorders. Orthop Clin North Am 27:679–709

28. Hansson GA, Balogh I, Bystrom JU, Ohlsson K, Nordander C, Asterland P, Sjolander S, Rylander L, Winkel J, Skerfving S, Malmo Shoulder-Neck Study Group (2001) Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands. Scand J Work Environ Health 27:30–40

29. Hart DL, Isernhagen SJ, Matheson LN (1993) Guidelines for functional capacity evaluation of people with medical conditions. J Orthop Sports Phys Ther 18:682–686

30. Hayes KW, Falconer J (1992) Reliability of hand-held dynamometry and its relationship with manual muscle testing in patients with osteoarthritis in the knee. J Orthop Sports Phys Ther 16:145–149

31. Holmstrom EB, Lindell J, Moritz U (1992) Low back and neck/shoulder pain in construction workers: occupational workload and psychosocial risk factors. Part 1: Relationship to low back pain. Spine (Phila Pa 1976) 17:663–671

32. Houtman IL, Goudswaard A, Dhondt S, van der Grinten MP, Hildebrandt VH, van der Poel EG (1998) Dutch monitor on stress and physical load: risk factors, consequences, and preventive action. Occup Environ Med 55:73–83

33. Innes E, Straker L (1999) Validity of work-related assessments. Work 13:125–152

34. Isernhagen Work Systems (1997) Functional capacity procedure manual 1st edn. Duluth, M.N.

35. Isernhagen SJ (1992) Return to work testing, functional capacity evaluation and work capacity evaluation. Orthopaedic Physical Therapy Clinics of North America 1:157–168

36. Isernhagen SJ (1988) Functional capacity evaluation. In: Isernhagen SJ (ed) Work injury management and prevention. Gaithersburg, MD: Aspen Publishers, pp 139–167

37. Johnston MV, Keith RA, Hinderer SR (1992) Measurement standards for interdisciplinary medical rehabilitation. Arch Phys Med Rehabil 73:S3–23

38. Jonsson B (1988) The static load component in muscle work. Eur J Appl Physiol Occup Physiol 57:305–310

39. Kamwendo K, Linton SJ, Moritz U (1991) Neck and shoulder disorders in medical secretaries. Part I. Pain prevalence and risk factors. Scand J Rehabil Med 23:127–133

40. Keyserling WM (2000) Workplace risk factors and occupational musculoskeletal disorders, Part 2: A review of biomechanical and psychophysical research on risk factors associated with upper extremity disorders. AIHAJ 61:231–243

41. Kilbom A, Horst D, Kemfert K, Richter A (1986) Observation methods for eduction of load and strain on the human body – A review. Abetarskyddsstyrelsen Publikation Service 171 (84):92

42. Kilbom A, Persson J (1987) Work technique and its consequences for musculoskeletal disorders. Ergonomics 30:273–279

43. King PM, Tuckwell N, Barrett TE (1998) A critical review of functional capacity evaluations. Phys Ther 78:852–866

44. Knave BG, Wibom RI, Voss M, Hedstrom LD, Bergqvist UO (1985) Work with video display terminals among office employees. I. Subjective symptoms and discomfort. Scand J Work Environ Health 11:457–466

45. Landis JR, Roth D, Straaton K (1991) Functional capacity evaluation in work disability. Work 1:37–47

46. Lau EM, Sham A, Wong KC (1996) The prevalence of and risk factors for neck pain in Hong Kong Chinese. J Public Health Med 18:396–399

47. Loeser JD, Melzack R (1999) Pain: an overview. Lancet 353:1607–1609

48. Malchaire J, Cock N, Vergracht S (2001) Review of the factors associated with musculoskeletal problems in epidemiological studies. Int Arch Occup Environ Health 74:79–90

49. Nachemson A, Vingard E (2000) Assessment of patients with neck and back pain: A best-evidence synthesis. In: Nachemson AL, Jonsson E (eds) Neck and Back Pain. Lippincott Williams & Wilkens, Philadelphia, pp 189–236

50. Norlander S, Aste-Norlander U, Nordgren B, Sahlstedt B (1996) Mobility in the cervico-thoracic motion segment: an indicative factor of musculo-skeletal neck-shoulder pain. Scand J Rehabil Med 28:183–192

51. Norlander S, Gustavsson BA, Lindell J, Nordgren B (1997) Reduced mobility in the cervico-thoracic motion segment--a risk factor for musculoskeletal neck-shoulder pain: a two-year prospective follow-up study. Scand J Rehabil Med 29:167–174

52. Ohlsson K, Attewell R, Skerfving S (1989) Self-reported symptoms in the neck and upper limbs of female assembly workers. Impact of length of employment, work pace, and selection. Scand J Work Environ Health 15:75–80

53. Ohlsson K, Attewell RG, Palsson B, Karlsson B, Balogh I, Johnsson B, Ahlm A, Skerfving S (1995) Repetitive industrial work and neck and upper limb disorders in females. Am J Ind Med 27:731–747

54. Onishi N, Nomura H, Sakai K, Yamamoto T, Hirayama K, Itani T (1976) Shoulder muscle tenderness and physical features of female industrial workers. J Hum Ergol (Tokyo) 5:87–102

55. Palmer KT, Walker-Bone K, Griffin MJ, Syddall H, Pannett B, Coggon D, Cooper C (2001) Prevalence and occupational associations of neck pain in the British population. Scand J Work Environ Health 27:49–56

56. Picavet HSJ, van Gils HWV, Schouten JSAG (2000) Musculoskeletal complaints in the Dutch population: Prevalences, consequences and risk groups (In Dutch: Klachten van het bewegingsapparaat in de Nederlandse bevolking: prevalenties, consequenties en risicogroepen).

57. Portney LG, Watkins MP (1993) Foundations of clinical research: Applications to practice. Appleton & Lange, Norwalk, Connecticut

58. Punnett L, Bergqvist U (1997) Visual Display Unit Work and Upper Extremity Musculoskeletal Disorders: A Review of Epidemiological Findings. Arbete och Hälsa 16:1–161

59. Reneman MF, Dijkstra PU, Westmaas M, Goeken LN (2002) Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. J Occup Rehabil 12:269–275

60. Reneman MF, Brouwer S, Meinema A, Dijkstra PU, Geertzen JH, Groothoff JW (2004) Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in healthy adults. J Occup Rehabil 14:295–305

61. Reneman MF, Soer R, Gerrits EH (2005) Basis for an FCE methodology for patients with work-related upper limb disorders. J Occup Rehabil 15:353–363

62. Rosecrance JC, Cook TM, Wadsworth CT (1992) Prevalence of musculoskeletal disorders and related job factors in 900 newspaper workers. In: Kumar S (ed) Advances in Industrial Ergonomics and Safety IV. Taylor & Francis, Philadelphia, pp 141–146

63. Rossignol AM, Morse EP, Summers VM, Pagnotto LD (1987) Video display terminal use and reported health symptoms among Massachusetts clerical workers. J Occup Med 29:112–118

64. Sakakibara H, Miyao M, Kondo T, Yamada S (1995) Overhead work and shoulder-neck pain in orchard farmers harvesting pears and apples. Ergonomics 38:700–706

65. Schibye B, Skov T, Ekner D, Christiansen JU, Sjogaard G (1995) Musculoskeletal symptoms among sewing machine operators. Scand J Work Environ Health 21:427–434

66. Skov T, Borg V, Orhede E (1996) Psychosocial and physical risk factors for musculoskeletal disorders of the neck, shoulders, and lower back in salespeople. Occup Environ Med 53:351–356

67. Soer R, Gerrits EH, Reneman MF (2006) Test-retest reliability of a WRULD functional capacity evaluation in healthy adults. Work 26:273–280

68. Spangfort E (1995) Clinical aspects of neck-and-shoulder pain. Scand J Rehabil Med Suppl 32:43–46

69. Thorn DW, Deitz JC (1989) Examining content validity through the use of content experts. The Occupational Therapy Journal of Research 9 (6):334–346

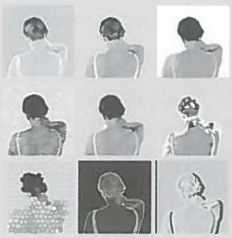
70. Viikari-Juntura E, Martikainen R, Luukkonen R, Mutanen P, Takala EP, Riihimaki H (2001) Longitudinal study on work related and individual risk factors affecting radiating neck pain. Occup Environ Med 58:345–352

71. Walker-Bone K, Cooper C (2005) Hard work never hurt anyone--or did it? A review of occupational associations with soft tissue musculoskeletal disorders of the neck and upper limb. Ann Rheum Dis 64:1112–1117

72. Weevers HJ, van der Beek AJ, Anema JR, van der Wal G, van Mechelen W (2005) Work-related disease in general practice: a systematic review. Fam Pract 22:197–204

73. Winkel J, Mathiassen SE (1994) Assessment of physical work load in epidemiologic studies: concepts, issues and operational considerations. Ergonomics 37:979–988

Chapter 8



General discussion

1. INTRODUCTION

This thesis focused on clinimetric properties of instruments to measure disability in patients with non-specific chronic neck pain (CNP). Three studies analyzed neck disability questionnaires, one study analyzed a general health questionnaire and one study analyzed physical examination as part of the clinical assessment in patients with non-specific CNP. In the sixth study a 'functional capacity evaluation' was presented which has content validity for determining functional capacity in patients with work-related neck disorders. The main research questions posed in the general introduction (Chapter 1) are answered in "Main findings". Methodological issues concerning these studies are considered. Clinical implications and recommendations for future research will be presented.

2. MAIN FINDINGS

In the first study (Chapter 2) the translation and cultural adaptation of the Neck Pain and Disability Scale (NPAD) was made following the forward and backward translation procedure [4, 61]. Test-retest reliability was acceptable for both NPAD and Neck Disability Index (NDI). NPAD showed larger 'instability' in the test-retest scores compared to the NDI.

In the second study (Chapter 3) the content validity, internal consistency and construct validity of NPAD and NDI were interpreted as good because 26 of the 27 predefined hypotheses were not rejected.

In the third study (Chapter 4) relevant changes in total scores on both NPAD and NDI after completion of a multidisciplinary rehabilitation program were assessed. This resulted in higher Minimal Detectable Change (MDC) than Minimal Important Change (MIC) for changes on NPAD and NDI and consequently with different amounts of certainty that the patient had improved. Furthermore this study demonstrated that the responsiveness of NPAD and NDI was similar and acceptable.

In the fourth study (Chapter 5) the construct validity of the Short Form 36 Health Survey (SF-36) was interpreted as good because 12 of the 16 predefined hypotheses were not rejected.

In the fifth study (Chapter 6) the overall interobserver reliability of Physical Dysfunction Severity (PDS) was acceptable, however, significant interobserver differences were obtained on PDS subscale 2 (limitation of passive cervical range of movement). The construct validity of PDS appeared satisfactory.

The overall interobserver reliability of the modified cervical non-organic signs (mcNOS) was acceptable, however the Kappa coefficient for interobserver agreement for the individual cervical nonorganic signs ranged from poor to acceptable. The construct validity of mcNOS appeared satisfactory.

In the sixth study (Chapter 7) a functional capacity evaluation (neck FCE) instrument was developed, which is content valid for determining functional capacity in patients with work-related neck disorders. The neck FCE was based on physical work-related risk factors for neck pain identified from epidemiologic literature. These risk factors were operationally defined into 8 tests of the neck FCE and the specific objectives of each test were described.

3. METHODOLOGICAL CONSIDERATIONS

3.1. Recruitment of patients

In this thesis patients with non-specific CNP in a tertiary rehabilitation center were studied. These patients were referred from general practitioners or medical specialists for diagnostic procedures as well as for advice and rehabilitation. About 33% of this group is admitted for multidisciplinary outpatient pain rehabilitation. To be admitted for a multidisciplinary pain rehabilitation, patients had to agree with the time-contingent approach to restore activities and to facilitate return to work.

Reasons for not having been admitted for pain rehabilitation were diverse; for instance patients were only reassured and or were advised how to cope with neck pain, patients were referred for further diagnostic procedures to other medical specialists, patients wanted primarily a pain-contingent approach or patients did not have treatment goals in return to work and/-or to restore in daily activities and participation. The characteristics of the patient group not-admitted for pain rehabilitation were not explored.

This procedure implied selection bias of patients in this thesis. Moreover the generalizability to other patient groups such as patients with neck pain in primary care setting or patients with a cervical root syndrome in secondary/tertiary setting may be limited. Patients with non-specific CNP admitted for multidisciplinary pain rehabilitation are younger and report more disability, longer duration of current episodes, more often trauma as cause of neck pain, more concomitant complaints, more previous treatments, and are more often involved in litigation and workers compensation claims, than patients with non-specific neck pain in primary care [26, 29, 30, 44, 47, 57, 58].

3.2. NPAD/NDI

In this thesis the measurement properties of NPAD and NDI divided over three domains viz. reproducibility, validity and responsiveness were analyzed.

3.2.1. Reproducibility

Reproducibility is the extent to which the same results are obtained on repeated tests when no real change in construct has occurred [13, 41]. Reproducibility contains two aspects: reliability and agreement [13]. Critical points in the evaluation of test-retest reproducibility are: were the patients "stable" (not changed in clinical status) in the retest interval on the construct neck pain related disability and was the time interval between the test sessions appropriate [36]. A (very) short time interval increases the chance that patients have not changed but increases the probability of carry over or recall effects. A time interval of about two weeks is often considered appropriate [36].

Reliability

As measure for reliability the intraclass correlation coefficient (ICC) was used. It assesses not only the strength of the correlation between two repeated measures but also if all measures on each subject are identical and do not differ systematically. From the studies included in this thesis (Chapter 2 and 4) it becomes clear that one value for reliability cannot be provided because previous studies reported a range of ICCs on NPAD (0.81 to 0.98) [1, 5, 32, 34, 63] and NDI (0.50 to 0.97) [1, 5, 8, 9, 32-35, 38, 55, 59, 63]. Perceived change in clinical status in the retest interval was assessed in only half of the abovementioned NPAD and NDI studies. Apart from that it seems not to have resulted in differences in the extent of ICCs. In general lower ICCs are found in studies with longer retest intervals as in other region specific questionnaires was reported [6]. Remarkably enough the study with the lowest ICC had a retest interval of only two and a half days [9]. This study was performed in 'stable' patients with mechanical neck pain (mean duration 82.4 days) in outpatient orthopedic physical therapy clinics. It is likely that other factors than the retest interval duration such as non-specific or specific neck pain, symptom duration (acute, sub-acute or chronic), patient setting (primary, secondary or tertiary care) and mean disability score on the questionnaires could also have influenced ICCs.

Agreement

To quantify the agreement, the test retest score stability over time on individual level the "limits of agreement" (LOA) were calculated (Chapter 2). Changes exceeding the LOA can be labeled as change beyond measurement error. Another approach to assess the measurement error is to calculate the minimal detectable change (MDC) (Chapter 4). In Chapter 2 the LOA of the NPAD was somewhat higher than in a previous study [63] but in that study the retest interval was one day. The LOA of the NDI was similar to most other NDI studies with shorter retest intervals [34, 35, 55, 59, 63]. As in other studies the NPAD showed a relatively higher LOA (larger measurement error) compared to the NDI [35, 55, 59, 63] which may be explained by differences in operational definitions of 'neck disability' between items of NPAD and NDI [56, 61]. Post-hoc analysis in the study presented in Chapter 2 showed that the amount of variation of the NPAD could not be attributed to individual items of the questionnaire.

In the study described in Chapter 4 the MDC of the NPAD was substantially higher than in a previous study [5]. This difference can be attributed to differences in the reported ICC [5]. The impact of the measured ICC for the MDC calculation was also obvious in our data of NPAD. With the results of Chapter 2 the MDC would be 22.4 and in Chapter 4 the MDC was 31.7. The MDC of the NDI in Chapter 4 falls in the range between 1.7 and 13.4 as reported in previous NDI studies [7–9, 42, 51, 59, 65, 66]. Apart from different patient populations, the observed differences are most likely the result of different formula for the MDC calculation (1.96 or 1.65x√2 xSEM) and large ranges in SEM (0.60 to 4.4) in these NDI studies [7–9, 42, 51, 59, 65, 66]. The relatively higher MDC for the NPAD than for the NDI in Chapter 4 may be explained by differences in the assessed ICCs on NPAD and NDI and by differences in operational definitions of neck disability between items of NPAD and NDI [56, 61]. Post-hoc analysis in the study presented in Chapter 4 showed that the amount of variation of the NPAD in stable patients at the end of the treatment could be attributed to significant differences in 7 individual items (2, 6, 8–12) of the NPAD.

In summary, the reliability of the NPAD is somewhat lower than of the NDI and the measurement error is relatively higher for the NPAD than the NDI.

3.2.2. Validity

Validity is the extent to which a questionnaire measures the construct it is supposed to measure. In this thesis 3 aspects of validity (content validity, internal consistency and construct validity) were tested.

Content validity

Content validity is the extent to which items of the questionnaire reflect all aspects of the construct to be measured. The construct in this study was the self-reported disability in patients with CNP. The International Classification of Functioning Disability and Health (ICF) defines disability as an umbrella term for impairments in bodily functions and structures, activity limitations and participation restrictions [52, 62]. Both NPAD and NDI include various combinations of impairment items, activity items and participation items [56, 61] (Table 8.1). So the construct self-reported disability is covered differently by the items of NPAD and NDI. This could be an explanation for the fact that in all studies, including ours, where both questionnaires were used the NPAD scores were approximately 10% higher than the NDI scores [32, 34, 38, 63].

The potential limitation of NPAD and NDI is that they use fixed questions that constrain the scope of the evaluation to the specific items included. Therefore the NPAD and NDI may include items not relevant to some patients and may not include items of importance to other patients. In contrast to the NPAD and NDI, the Problem Elicitation Technique (PET) [17, 25], the 'Patient-Specific Functional Scale' (PSFS) [60] and the Canadian Occupational Performance Measure [16, 18] are interviewer administered instruments that ask the patients to identify their own limitations/problems that have risen as a result of the neck pain. Of the most commonly

Table 8.1 Item profile of NPAD and NDI according to the International Classification of Functioning Disability and Health [62]

Body functions and body structures			Activities			Participation		
Items	NPAD	NDI	Items	NPAD	NDI	Items	NPAD	NDI
Pain intensity	X	X	Lifting	–	X	Personal relationships	X	–
Pain on average	X	–	Reading	–	X	Recreational activities	X	X
Pain at worst	X	–	Driving/riding	X	X	Social activities	X	–
Effect of pain pills	X	–	Personal care	X	X	Working	X	X
Headaches	–	X	Standing	X	–			
Thinking/concentration	X	X	Sleeping	X	X			
Pain changed outlook on life and future	X	–	Walking	X	–			
Pain affect emotions	X	–	Working overhead	X	–			
Neck stiffness	X	–						
Turning neck	X	–						
Looking up/down	X	–						

X, item included; –, item not included.

mentioned problems in the aforementioned PET-studies in both NPAD and NDI 'fatigued during the day' and 'frustration' were not included. The concomitant complaints profile of our study population makes also clear that neither the NPAD nor the NDI captures the full spectrum of limitations/disabilities of the patients with non-specific CNP.

The distribution of the total scores and the completeness of item responses were similar to other studies [5, 34, 37, 48, 63]. In our study floor and ceiling effects were <13% for all NPAD-items, but in studies with samples with lower disability levels floor effects of 6 to 14 items were found [5, 34]. For the NDI, the items 'personal care' and 'sleeping' had floor effects as was also reported in previous studies [22, 25, 34, 59].

Internal consistency

Internal consistency is the extent to which all items measure the same construct [45]. It was assessed with Cronbach's alpha for the total scales and standardized item total score correlations of NPAD and NDI. The NPAD is claimed to be a questionnaire with 4 underlying dimensions [61]. Factor analyses in other language NPADs identified 1 to 4 factors on which different items were loading [5, 10, 37, 38, 48, 63]. We calculated a single Cronbach's alpha for NPAD and NDI because their factor structure (1, 2, 3 or 4 factors for NPAD and 1 or 2 factors for NDI) is unclear and because in the original English versions single Cronbach's alphas for the total scales were calculated [56, 61].

Construct validity

Construct validity refers to the extent to which the scores of the questionnaire correlate as hypothesized with other assessment tools or constructs [36]. To test the construct validity comparisons were made with other constructs known to be associated with neck pain, neck pain related disability or generic health. NPAD and NDI were at least fairly correlated with all 8 SF-36 domains, and the strength of these correlations were in most of the SF-36 domains similar for NPAD and NDI. As hypothesized the correlation between NPAD and VAS_{pain} was slightly higher than for NDI and VAS_{pain} and this was similar to other studies [32, 63]. As expected the correlation between NPAD and NDI and the $VAS_{disability}$ was similar to that of other studies [32, 63]. NPAD and NDI are multi-item instruments and time consuming to use. For that reason it may be attractive to use a single item measure such as $VAS_{disability}$ with the intention to measure the whole construct at once.

The correlation between NPAD and NDI was 0.77, similar to other studies (0.66–0.86) This suggests that these questionnaires measure similar constructs [3, 17, 38, 63], although only 7 items of both questionnaires covered similar problems/limitations. In a post-hoc analysis the strength of the relationship between scores of similar NPAD and NDI items ranged from 0.50 to 0.72 (Table 8.2). All NPAD item scores were higher than similar items of the NDI. Probably the 0 to 5 VAS of the NPAD items leads to different results in comparison with the 6 different assertions

Table 8.2 Mean (SD) scores and Spearman correlations for similar items of the NPAD–DLV and NDI–DLV (n = 112) [30]

Items	NPAD mean (SD) (range score 0–5)	NDI mean (SD) (range score 0–5)	Correlation
Pain intensity	2.6 (1.0)	2.2 (0.8)	.72
Sleeping	2.5 (1.5)	1.9 (1.3)	.66
Thinking/concentration	2.7 (1.5)	2.0 (1.3)	.65
Driving/riding	2.6 (1.3)	2.4 (1.2)	.64
Personal care	1.7 (1.3)	0.7 (0.8)	.70
Recreational activities	3.0 (1.1)	2.4 (1.0)	.50
Working	3.4 (1.1)	2.5 (1.2)	.58

construction of the NDI with a score between 0 and 5 points. For example in the 'recreation' item the NPAD question is 'does your pain interfere with recreational activities' (anchors: 'not at all' and 'always'). In the NDI item the assertions vary from 'no pain at all', 'some pain in all recreation activities' and 4 assertions in relation to the possibility to engage in respectively 'all', 'a few', 'hardly any' and 'any recreation activities' because of neck pain.

Summary

The construct self-reported neck pain related disability is covered differently by the items of NPAD and NDI and both questionnaires do not capture the full spectrum of disability of the patients with non-specific CNP. Floor effects of certain NPAD and NDI items are expected in populations with lower disability levels. Factor analysis in previous NPAD and NDI studies identified a different factor structure. For the future we suggest to perform a factor analysis for the NPAD and NDI items in accordance with the ICF model. Construct validity of NPAD and NDI was largely similar when they were compared with other constructs known to be associated with neck pain, neck pain related disability of generic health.

3.2.3. Responsiveness

Responsiveness is the ability of a questionnaire to detect change over time in the construct to be measured [36]. In this thesis (Chapter 4) the responsiveness of respectively NPAD and NDI was compared using correlations between change scores of NPAD and NDI and global perceived effect (GPE) scores and by using the Areas Under the receiver operator characteristic Curves (AUC).

The correlations between change scores and GPE were similar to those of most NPAD (range 0.42–0.59) [21, 34, 64] and NDI (range 0.19–0.58) studies [8, 9, 34, 49, 65, 66]. The NPAD and NDI

both had an AUC of 0.75 which was a satisfactory result and in line with results found by other studies (range 0.57–0.90) [8, 9, 34, 42, 49, 51, 65, 66]. In one study [34] in which responsiveness of NPAD and NDI was also compared using GPE as external criterion an AUC of 0.79 for both was reported. Remarkably the largest AUC of 0.90 for the NDI was reported in a study using a prognostic estimate of change as external criterion made by clinicians at a patient's initial visit [51].

To evaluate the effect of treatment programs in patients with neck pain there is a need to define minimum changes in scores on questionnaires that are relevant from patients', clinicians' or socioeconomic perspectives [53]. The degree to which one can assign qualitative meaning to quantitative (change) scores is an aspect of interpretability [36]. In this thesis two concepts of interpretability were used. In the distribution-based approach the LOA (Chapter 2) and the MDC (Chapter 4) were used as calculations of the measurement error and were discussed earlier in this Chapter.

The anchor-based approach assessed, which minimal important change (MIC) on NPAD or NDI corresponded with an imported change on the self-reported global perceived effect (GPE) by the patient at the end of the treatment program (Chapter 4). MIC is defined as "the smallest change that is important to patients" [11, 12, 27, 51, 53]. How to classify the smallest important change and how to classify patients as improved or stable with GPE scale levels, is an arbitrary decision [7–9, 12, 42, 59, 64–66]. In most studies using GPE as external standard a 15 point scale was used with ≥3 (moderately better) as cut-off to distinguish improved from stable patients [8, 9, 42, 65, 66]. In this thesis patients were classified as improved when they scored completely recovered or much recovered to reflect important improvement similar to other studies [14, 42]. Consequently, this may lead to overestimation of the MIC.

In the study presented in Chapter 4 the MICs for NPAD and NDI were respectively 11.5 and 3.5. No values of MIC for NPAD have been reported by others. MIC for NDI has been reported to range from 3.5 to 9.5 [7–9, 42, 51, 65, 66]. Differences between these studies and the present study could be the result of different external criteria (prognostic estimate of change [51], Health Transition Item of SF-36 [7], GPE by patient [8, 9, 42] or by patient and therapist [65]), the number of scale levels of the external criteria, the combination of scale levels to form the improved and stable group, characteristics of population (such as age, nature and acuity of neck condition, patient setting, baseline scores), treatment and period of follow up [7–9, 42, 65, 66]. Smaller values for MIC compared with MDC or LOA were observed in almost all neck pain studies including this thesis [7–9, 42, 65, 66]. In other words a part of the improved patients according the MIC cut-offs falls within the measurement error of the questionnaires.

The most frequently used external criterion in neck pain studies is the GPE. The advantage in clinical practice is that the GPE reflects the patient's personal assessment of overall improvement instead of only one aspect of impairment (e.g. pain or function) [47]. Nevertheless the use of the GPE was also criticized because it consists of only one question and a patient's ability to recall his/her previous health status is questionable [18, 36]. Moreover the retrospective rating

with the GPE is more likely to correlate with the patient's status at follow up than with status at baseline [36]. This challenges the accurate representation of true change between baseline and follow up. As an alternative to the GPE it is suggested to use the change score for a patient's global rating of disability assessed at baseline and follow up [47, 48]. In neck pain studies the correlation between $VAS_{disability}$ with NPAD and NDI ranged respectively from 0.57 to 0.69 and from 0.50 to 0.71 [30, 32, 63]. In a post-hoc analysis in Chapter 4 the correlations of change scores of NPAD and NDI with $VAS_{disability}$-change were 0.71 (95% CI 0.58 to 0.80) and 0.53 (95% CI 0.35 to 0.67) which, especially for the NPAD, is a much stronger correlation than between change scores of NPAD and NDI and GPE. Therefore the $VAS_{disability}$-change could possibly be preferred as self-reported external criterion.

In a more methodological reflection it was stressed that in the assessment of MIC the external criterion should be an independent objective measure of the same construct and not some other related self-report measure as the GPE [19, 20]. As independent non-self-reported external criterion 1-year post-treatment socioeconomic outcomes (return-to-work, work retention, percentage seeking treatment from a new provider, and mean number of visits to the new provider) were reported in patients with chronic disabling occupational spinal disorders [19, 20]. The result of the above mentioned study demonstrated that the consensus-based decision of a 30% reduction from baseline [39] on the Oswestry Disability Index and the SF-36 were poorly associated with the socioeconomic outcomes [19]. This challenges the use of the GPE as external criterion. It is suggested therefore in the post treatment and 1-year post treatment evaluation, also to use the $VAS_{disability}$-change and the above mentioned socioeconomic outcomes.

Because of the outcome of the third study (Chapter 4) it is of interest with to compare different instruments to measure improvement or recovery. Looking at group means in NPAD and NDI scores at baseline and follow up there is a significant small to substantial improvement for all 'stable' and 'improved' patients. The importance of establishing cut off values for MDC and MIC lies in the interpretation of relevant change in the individual patient in contrast to the group means of change. When the percentage of patients achieving the cut off value for MDC is calculated with the NPAD 16% and with the NDI 23% show statistically relevant improvement. When the percentage of patients achieving the cut off value for MIC is calculated with the NPAD 56% and with the NID 57% has recovered clinically relevant change. With the GPE 59% of the patients has completely or much recovered. As a rule of thumb in patients with low back pain a 30% change from baseline may be considered clinically important change for pain and functional status [39]. If this criterion would be applied with the NPAD 51% and with the NDI 38% of the patients would have improved clinically relevant.

The differences between patients with non-specific neck pain in 3 controlled trials in a Dutch primary care setting and patients in a tertiary rehabilitation setting in relation to self-reported disability levels and perceived recovery at the end of the intervention period are presented

Table 8.3 Baseline characteristics, self-reported disability levels and perceived recovery in patients with non-specific neck pain of controlled trials in a primary care setting and patients in a tertiary rehabilitation setting

	Hoving et al., 2006 [26]			Pool et al., 2010 [44]		Vonk et al., 2009 [57]		Jorritsma et al., 2012 [31]
	MT n = 60	PT n = 59	GP n = 64	MT n = 75	BGA n = 71	PT n = 70	BGA n = 68	MPR n = 111
Age (mean SD)	44.6(12.4)	45.9(11.9)	45.9(10.5)	43.6(11.1)	44.5(12.0)	45.7(12.3)	45.7(12.1)	38.7 (11.4)
Female (%)	57	70	56	68	59	61	63	63
Duration of neck pain (%)								
2–6 weeks	48	46	50	0	0	0	0	0
4–12 weeks	n.r.	n.r.	n.r.	100	100	0	0	0
7–12 week	22	25	31	n.r.	n.r	0	0	0
≥ 13 weeks	30	29	19	0	0	100	100	100
Intervention period	6 wk	6 wk	6 wk	6 wk	6 wk	9 wk	9 wk	3–4 mo
Mean number of visits	6	9	2	5.2	8.2	11.2	6.6	65–85
Drop outs	1 (2%)	0 (0%)	0 (0%)	3 (4%)	3 (4%)	1 (1%)	2 (3%)	35 (32%)
NDI (scale 0–50)								
Baseline	13.6	13.9	15.9	13.4	14.7	15.4	15.2	21.6
Post treatment	6.4	9.4	10.6	6.3	5.6	13.2	11.0	15.3*
Recovered (%)	69	51	36	68	61	42	40	41

MT, manual therapy; PT, physiotherapy; GP, general practitioner; BGA, behavioral graded activity by PT; NDI, Neck Disability Index; MPR, multidisciplinary pain rehabilitation; SD, standard deviation; n.r., not reported; wk, weeks; mo, months. Recovered: completely recovered and much improved with global perceived effect (GPE) assessment.

* Patients who completed MPR.

in Table 8.3 [26, 31, 44, 57]. Patients with CNP in primary and tertiary care settings do have higher disability levels and lower percentages of recovering with GPE assessment at the follow up at the end of the intervention period than patients with acute or sub-acute neck pain in a primary care setting. The percentage of dropouts is much higher in a tertiary care setting than in a primary care setting.

Summary
Responsiveness was similar in both NPAD and NDI. Relevant changes on both NPAD and NDI assessed with MDC (or LOA) and MIC resulted as in almost all neck pain studies in higher cut-offs for change scores for MDC. Including the values for MDC and MIC presented in this thesis a broad range of these values as result of partly unknown factors has been reported in previous studies.

3.3. SF-36

For the assessment of the construct validity of the SF-36 in patients with non-specific CNP, there is no established gold standard of general health to compare the SF-36 with. Overall, the SF-36 domain scores in the present study make clear that CNP has great impact on general health perception. This was in line with three other studies [23, 28, 50] in physical therapy outpatients, a mixture of rehabilitation clinic inpatients and physical therapy outpatients, and in spine care center patients. The domain score profiles in the aforementioned studies and in the present study were comparable.

As expected scores on PCS and MCS in our setting were significantly lower than those in patients with non-specific neck pain in Dutch primary care setting [43]. As in other studies [24, 35, 43, 46] the PCS score in the present study was lower than the MCS score. Apparently, in patients with neck pain physical health is affected to a greater extent than mental health. It is very likely that this is the result of the great impact of the domain scores 'physical functioning', 'role physical' and 'bodily pain' on the PCS score. A further indication that physical health is effected to a greater extent is that higher correlations of PCS than of MCS with neck disability were reported [24, 35, 46].

Because of the clinical characteristics of the patients in our tertiary rehabilitation setting we expected, as in patients with back pain, significantly lower PCS and MCS scores in the litigation group and in the workers compensation group [15, 30]. In the present study 33% of the patients were in litigation and although their PCS and MCS scores were somewhat lower than of non-litigants, the differences were not significant. Because of the outcome of our second study we suppose that the contrast between groups with and without litigation is larger when using a more specific neck pain related disability questionnaire than using the generic SF-36 [30]. In the present study 57% of the patients were receiving workers compensation and their PCS and

MCS scores were lower than of patients who were not receiving workers compensation; for the PCS scores this difference was significant. Because physical health in patients with neck pain is more affected than mental health, it is very likely that this has greater impact on PCS scores than on MCS scores in patients with workers compensation.

3.4. PDS and mcNOS

Small but significant differences were observed between the first and second examiner in PDS and mcNOS. All our patients were reexamined after a waiting period (mean 20 days) and it was deemed unlikely that they changed in clinical status because in patients with a similar retest interval no substantial change in neck pain related disability was observed. Although the construct physical impairment and behavioral signs are different from self-reported disability, we assume that the influence of this interval is not substantial. Besides other explanations as discussed in Chapter 6 we suppose that possibly the observer differences may have influenced the outcomes of PDS and mcNOS. Differences in techniques between the examiners during the physical examination and differences in interpretation may possibly have influenced the outcomes. We were reluctant to create an `ideal observers world` which could hardly be generalized to a ´real world´ situation in clinical practice. Therefore the training phase was relatively brief. To optimize the interobserver reliability of PDS and mcNOS a more intensive training in the protocol of the physical examination of PDS and mcNOS is desirable.

With respect to the PDS only subscale 2 showed significant differences. For the assessment of subscale 2 of the PDS we recommend to measure flexion, extension and rotations in a lower position of the examination table. To measure the lateral flexion a higher position of the examination table is recommended. This makes the visual estimation of the passive range of motions of these movements easier.

With respect to the mcNOS we suggest to standardize the amount of pressure with the deep tenderness test and the axial loading simulation test to respectively 2 to 3 pounds and 3 pounds. Additionally in the axial loading and simulated rotation test we suggest to score these tests positively when neck pain is produced instead of (more) neck pain is produced. Furthermore we suggest to use as ´motor loss´ tests the isometric strength tests of the abductors, lateral and medial rotators of the shoulders, the elevators and upward rotators of the scapulas and the flexors and extensors of the elbows. Additionally, a better defining of overreaction tests is desirable.

We propose to repeat the reproducibility study of the mcNOS in accordance with the protocol presented by the International Federation of Manual/Musculoskeletal Medicine (FIMM) with a training phase, an overall agreement phase (>80%) and a study phase with a (mcNOS ≥3) case prevalence of 50% [40].

3.5. FCE

For the development of a neck FCE, physical work-related risk factors that might have contributed to the development of work-related neck disorders (WRND) were identified from reviews of epidemiologic literature of WRND. An important limitation of the literature was the overall absence of a definition of risk factors. The nature of physical risk factors was indeed often described in terms of repetitive and/or forceful movement, awkward positions and static contractions but a more precisely description of the content of risk factors was only reported in one review [2].

3.6. Strengths and limitations

Strength of the three NPAD/NDI studies are that the clinimetric properties are compared head to head with explicit pre-defined hypotheses testing of the validity. A strength of the SF-36 study is that the construct validity is tested using explicit pre-defined hypotheses. For use in daily clinical practice a simple instrument -the PDS- is modified and operationalized for the estimation of the physical impairment in patients with neck pain. Moreover a modified cervical NOS is constructed and for the first time the prevalence of the non-organic signs and the interobservers reliability of the total scores of mcNOS are reported. A neck-FCE was developed based on the present understanding of the physical factors for patients with work-related neck disorders.

In the first five studies patients with non-specific CNP admitted for an outpatient tertiary rehabilitation treatment were included in a sample largely consisting of patients with moderate neck pain and disability. So generalizability of interobserver reliability, test-retest reproducibility, validity and responsiveness to populations outside this setting may be limited. The hypotheses and the cut-off points that were used in the validity studies of NPAD/NDI and SF-36 study were based on previous studies without a systematic appraisal of the quality of these studies. In the study regarding interobserver reliability and construct validity of PDS and mcNOS the test-retest interval was relatively long and the change in clinical status was notchecked upon. Therefore the interobserver reliability was probably underestimated. Moreover the sample size in this study was adequate for its purpose, although a larger sample is preferred for a validity study.

In the sixth study the selection of physical risk factors in the literature was based on previous reviews without a methodical and qualitative analysis of the validity of these reviews. Furthermore, we do not know if certain physical work-related activities are the primary factors for the development of WRND or if neck pain in patient's is more inconvenient when performing these activities. Physical risk factors could be easily operationalised into 8 tests of the neck FCE. However the question remains if all relevant physical risk factors that might have contributed to the development of the WRND are covered by the neck FCE for the individual WRND patient with a particular occupation.

4. CLINICAL IMPLICATIONS FOR MEASURING NECK PAIN RELATED DISABILITY

Both NPAD and NDI have acceptable measurement properties but the NDI is slightly better. An advantage of the NPAD is, however, a broader item profile according to the ICF. Both questionnaires could be used in patients with neck pain in different settings (primary, secondary or tertiary care) and with different symptom duration (acute, sub-acute and chronic) to measure self-reported neck pain related disability and treatment efficacy. When measuring changes in total scores on NPAD and NDI after a treatment program, clinicians should be aware that choosing either the MDC, LOA or the MIC cut-off will lead to different values and amounts of certainty to which the observed change is relevant. Applying the more conservative MDC or LOA, the certainty that the change score is relevant and larger than the measurement error, is high. The amount of certainty needed may depend on the consequences in patient care and this could be a case by case decision.

The SF-36 provides a more comprehensive measure of self-reported disability than the NPAD and NDI, although the construct neck pain related disability is reflected better by the NPAD and NDI. In addition the advantage of using the SF-36 is the possibility to compare the patients with non-specific CNP in this thesis with other types of disorders. The SF-36 could be used in patients with neck pain in different settings and with different symptom duration to measure self-reported disability.

The PDS should be considered as measure of physical impairment of the cervical spine in patients with neck pain in different settings and as outcome measure in primary care. The mcNOS should be considered as a clinical diagnostic screening tool in patients with CNP particularly in a secondary or tertiary setting. Assessment of both the mcNOS and physical impairment of cervical spine during physical examination may offer relevant information in the biopsychosocial diagnosis of patients with CNP.

Based on the assumed content validity, the neck FCE may be a performance based assessment to measure a patient's disability to perform work-related activities in relation to the work context in secondary and tertiary care settings. The outcome of a neck FCE can serve several medical disciplines such as rehabilitation and occupational medicine in diagnostic procedures, treatment plan and as outcome measure. Before the neck FCE can be used for these goals, other psychometric properties such as reliability should be studied.

5. RECOMMENDATIONS FOR FURTHER RESEARCH

Findings of the presented NPAD/NDI studies in this thesis underline the need of further research, especially of the reproducibility, responsiveness and interpretability of the NPAD in other patient groups (e.g. acute, sub-acute, primary care and cervical root syndrome, post-neck

surgery). Furthermore, it would be of interest to examine the internal consistency of NPAD and NDI in accordance with the ICF framework with the items divided in body functions and body structures, activities and participation.

For future studies of responsiveness and interpretability of NPAD and NDI it is recommended, in addition to the GPE as external criterion, to use also the $VAS_{disability}$-change, socioeconomic outcomes, prognostic estimate of expert based change made by clinicians and a performance based outcome as external criteria for relevant change. It is very likely that the types of external criteria used depend on the patient setting [19].

Findings of this study underline the need of further research with the SF-36 to assess the validity in other patient groups with neck pain and to assess the responsiveness and interpretability in comparison with different external criteria for relevant change.

Further studies are warranted to develop PDS and mcNOS and to investigate the inter- and intraobserver reliability of PDS and mcNOS. To assess the validity of PDS and mcNOS studies with larger sample sizes arerecommanded. Moreover, to test the usefulness of mcNOS as screening tool for patients who require more detailed psychosocial assessment, an in depth analysis of relationships between mcNOS and psychological and social factors is needed.

Recently, a study has been published in which the test-retest reliability of an FCE for patients with a whiplash-associated disorder (WAD) was assessed [54]. In this WAD-FCE 5 of the neck FCE tests (repetitive left and right side reaching, static overhead work, front carry, overhead lift) were integrated. These neck FCE tests showed good to excellent test-retest reliability [54]. Further research with the proposed neck FCE is necessary to analyze other measurement properties.

REFERENCES

1. Agarwal S, Allison GT, Agarwal A, Singer KP (2006) Reliability and validity of the Hindi version of the Neck Pain and Disability Scale in cervical radiculopathy patients. Disabil Rehabil 28:1405–1411
2. Ariens GA, van Mechelen W, Bongers PM, Bouter LM, van der Wal G (2000) Physical risk factors for neck pain. Scand J Work Environ Health 26:7–19
3. Aslan E, Karaduman A, Yakut Y, Aras B, Simsek IE, Yagly N (2008) The cultural adaptation, reliability and validity of neck disability index in patients with neck pain: a Turkish version study. Spine (Phila Pa 1976) 33:E362–E365
4. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25:3186–3191
5. Bremerich FH, Grob D, Dvorak J, Mannion AF (2008) The Neck Pain and Disability Scale: cross-cultural adaptation into German and evaluation of its psychometric properties in chronic neck pain and C1-2 fusion patients. Spine 33:1018–1027

6. Brouwer S, Kuijer W, Dijkstra PU, Goeken LN, Groothoff JW, Geertzen JH (2004) Reliability and stability of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. Disabil Rehabil 26:162–165

7. Carreon LY, Glassman SD, Campbell MJ, Anderson PA (2010) Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J 10:469–474

8. Cleland JA, Fritz JM, Whitman JM, Palmer JA (2006) The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine (Phila Pa 1976) 31:598–602

9. Cleland JA, Childs JD, Whitman JM (2008) Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. Arch Phys Med Rehabil 89:69–74

10. Cook C, Richardson JK, Braga L, Menezes A, Soler X, Kume P, Zaninelli M, Socolows F, Pietrobon R (2006) Cross-cultural adaptation and validation of the Brazilian Portuguese version of the Neck Disability Index and Neck Pain and Disability Scale. Spine 31:1621–1627

11. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY (2008) Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J 8:968–974

12. Crosby RD, Kolotkin RL, Williams GR (2003) Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 56:395–407

13. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ (2001) Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 17:479–487

14. de Vet HC, Ostelo RW, Terwee CB, van der RN, Knol DL, Beckerman H, Boers M, Bouter LM (2007) Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 16:131–142

15. Deyo RA (2000) Pain and public policy. N Engl J Med 342:1211–1213

16. Donnelly C, Carswell A (2002) Individualized outcome measures: a review of the literature. Can J Occup Ther 69:84–94

17. En MC, Clair DA, Edmondston SJ (2009) Validity of the Neck Disability Index and Neck Pain and Disability Scale for measuring disability associated with chronic, non-traumatic neck pain. Man Ther 14:433–438

18. Eyssen IC, Beelen A, Dedding C, Cardol M, Dekker J (2005) The reproducibility of the Canadian Occupational Performance Measure. Clin Rehabil 19:888–894

19. Gatchel RJ, Mayer TG (2010) Testing minimal clinically important difference: consensus or conundrum? Spine J 10:321–327

20. Gatchel RJ, Mayer TG (2010) Testing minimal clinically important difference: additional comments and scientific reality testing. Spine J 10:330–332

21. Goolkasian P, Wheeler AH, Gretz SS (2002) The neck pain and disability scale: test-retest reliability and construct validity. Clin J Pain 18:245–250

22. Hains F, Waalen J, Mior S (1998) Psychometric properties of the neck disability index. J Manipulative Physiol Ther 21:75–80

23. Hee HT, III Whitecloud TS,3rd, Myers L, Roesch W, Ricciardi JE (2002) Do worker's compensation patients with neck pain have lower SF-36 scores? Eur Spine J 11:375–381

24. Hermann KM, Reese CS (2001) Relationships among selected measures of impairment, functional limitation, and disability in patients with cervical spine disorders. Phys Ther 81:903–914

25. Hoving JL, O'Leary EF, Niere KR, Green S, Buchbinder R (2003) Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. Pain 102:273–281

26. Hoving JL, de Vet HC, Koes BW, Mameren H, Deville WL, van der Windt DA, Assendelft WJ, Pool JJ, Scholten RJ, Korthals-de Bos IB, Bouter LM (2006) Manual therapy, physical therapy, or continued care by the general practitioner for patients with neck pain: long-term results from a pragmatic randomized clinical trial. Clin J Pain 22:370–377

27. Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 10:407–415

28. Jette DU, Jette AM (1996) Physical therapy and health outcomes in patients with spinal impairments. Phys Ther 76:930–941

29. Jorritsma W, Reneman MF, Dijkstra PU, de Vries GE, Geertzen JHB The construct validity of the Short Form-36 Health Survey for patients with non-specific chronic neck pain. Submitted

30. Jorritsma W, de Vries GE, Dijkstra PU, Geertzen JH, Reneman MF (2012) Neck Pain and Disability Scale and Neck Disability Index: validity of Dutch language versions. Eur Spine J 21:93–100

31. Jorritsma W, Dijkstra PU, de Vries GE, Geertzen JH, Reneman MF (2012) Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index. Eur Spine J 21:2550–2557

32. Kose G, Hepguler S, Atamaz F, Oder G (2007) A comparison of four disability scales for Turkish patients with neck pain. J Rehabil Med 39:358–362

33. Kovacs FM, Bago J, Royuela A, Seco J, Gimenez S, Muriel A, Abraira V, Martin JL, Pena JL, Gestoso M, Mufraggi N, Nunez M, Corcoll J, Gomez-Ochoa I, Ramirez MJ, Calvo E, Castillo MD, Marti D, Fuster S, Fernandez C, Gimeno N, Carballo A, Milan A, Vazquez D, Canellas M, Blanco R, Brieva P, Rueda MT, Alvarez L, Del Real MT, Ayerbe J, Gonzalez L, Ginel L, Ortega M, Bernal M, Bolado G, Vidal A, Ausin A, Ramon D, Mir MA, Tomas M, Zamora J, Cano A (2008) Psychometric characteristics of the Spanish version of instruments to measure neck pain disability. BMC Musculoskelet Disord 9:42

34. Lee H, Nicholson LL, Adams RD, Maher CG, Halaki M, Bae SS (2006) Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. Spine 31:1841–1845

35. McCarthy MJ, Grevitt MP, Silcocks P, Hobbs G (2007) The reliability of the Vernon and Mior neck disability index, and its validity compared with the short form-36 health survey questionnaire. Eur Spine J 16:2111–2117

36. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. BMC Med Res Methodol 10:22

37. Monticone M, Baiardi P, Nido N, Righini C, Tomba A, Giovanazzi E (2008) Development of the Italian version of the Neck Pain and Disability Scale, NPDS-I: cross-cultural adaptation, reliability, and validity. Spine 33:E429–E434

38. Mousavi SJ, Parnianpour M, Montazeri A, Mehdian H, Karimi A, Abedi M, Ashtiani AA, Mobini B, Hadian MR (2007) Translation and validation study of the Iranian versions of the Neck Disability Index and the Neck Pain and Disability Scale. Spine 32:E825–E831

39. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC (2008) Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 33:90–94

40. Patijn J, Remvig L (2007) Reproducibility And Validity 2007-Protocol Formats For Diagnostics Procedures In Manual/Musculoskeletal Medicine.

41. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF (2002) Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine 27:515–522

42. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC (2007) Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine (Phila Pa 1976) 32:3047–3051

43. Pool JJ, Ostelo RW, Knol D, Bouter LM, de Vet HC (2010) Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain? Man Ther 15:111–116

44. Pool JJ, Ostelo RW, Knol DL, Vlaeyen JW, Bouter LM, de Vet HC (2010) Is a behavioral graded activity program more effective than manual therapy in patients with subacute neck pain? Results of a randomized clinical trial. Spine (Phila Pa 1976) 35:1017–1024

45. Portney LG, Watkins MP (2000) Foundations of Clinical Research 2nd Edition Prentice Hall Health. , Upper Saddle River, New Yersey 07458

46. Riddle DL, Stratford PW (1998) Use of generic versus region-specific functional status measures on patients with cervical spine disorders. Phys Ther 78:951–963

47. Schellingerhout JM, Heymans MW, Verhagen AP, Lewis M, de Vet HC, Koes BW (2010) Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. Spine (Phila Pa 1976) 35:E827–35

48. Scherer M, Blozik E, Himmel W, Laptinskaya D, Kochen MM, Herrmann-Lingen C (2008) Psychometric properties of a German version of the neck pain and disability scale. Eur Spine J 17:922–929

49. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M (2007) Responsiveness of pain and disability measures for chronic whiplash. Spine (Phila Pa 1976) 32:580–585

50. Stoll T, Huber E, Bachmann S, Baumeler HR, Mariacher S, Rutz M, Schneider W, Spring H, Aeschlimann A, Stucki G, Steiner W (2004) Validity and sensitivity to change of the NASS questionnaire for patients with cervical spine disorders. Spine 29:2851–2855

51. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B (1999) Using the Neck Disability Index to Make Decisions Concerning Individual Patients. 107–119

52. Stucki G (2005) International Classification of Functioning, Disability, and Health (ICF): a promising framework and classification for rehabilitation medicine. Am J Phys Med Rehabil 84:733–740

53. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34–42

54. Trippolini MA, Reneman MF, Jansen B, Dijkstra PU, Geertzen JH (2012) Reliability and Safety of Functional Capacity Evaluation in Patients with Whiplash Associated Disorders. J Occup Rehabil

55. Trouli MN, Vernon HT, Kakavelakis KN, Antonopoulou MD, Paganas AN, Lionis CD (2008) Translation of the Neck Disability Index and validation of the Greek version in a sample of neck pain patients. BMC Musculoskelet Disord 9:106

56. Vernon H, Mior S (1991) The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 14:409–415

57. Vonk F, Verhagen AP, Twisk JW, Koke AJ, Luiten MW, Koes BW (2009) Effectiveness of a behaviour graded activity program versus conventional exercise for chronic neck pain patients. Eur J Pain 13:533–541

58. Vos C, Verhagen A, Passchier J, Koes B (2007) Management of acute neck pain in general practice: a prospective study. Br J Gen Pract 57:23–28

59. Vos CJ, Verhagen AP, Koes BW (2006) Reliability and responsiveness of the Dutch version of the Neck Disability Index in patients with acute neck pain in general practice. Eur Spine J 15:1729–1736

60. Westaway MD, Stratford PW, Binkley JM (1998) The patient-specific functional scale: validation of its use in persons with neck dysfunction. J Orthop Sports Phys Ther 27:331–338

61. Wheeler AH, Goolkasian P, Baird AC, Darden BV (1999) Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. Spine 24:1290–1294

62. WHO (2001) ICF: International Classification of Functioning, Disability, and Health. World Health Organization, Geneva

63. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel M (2002) French translation and validation of 3 functional disability scales for neck pain. Arch Phys Med Rehabil 83:376–382

64. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel M (2004) The ability to change of three questionnaires for neck pain. Joint Bone Spine 71:317–326

65. Young BA, Walker MJ, Strunce JB, Boyles RE, Whitman JM, Childs JD (2009) Responsiveness of the Neck Disability Index in patients with mechanical neck disorders. Spine J 9:802–808

66. Young IA, Cleland JA, Michener LA, Brown C (2010) Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. Am J Phys Med Rehabil 89:831–839

Summary

Neck pain is one of the most common musculoskeletal disorders. The 1-year prevalence of chronic (>3 month duration) neck pain (CNP) in the general population is about 20%. In the large majority of patients with neck pain no specific underlying pathology can be established and the neck pain is labelled as "non-specific" or "mechanical" because it is produced or aggravated by neck movements or sustained neck postures. Limitations in activities and restrictions in participation in dailylife activities, including work, may be the result. The majority of total costs of neck pain in The Netherlands were costs due to sick leave and disability payment.

Disability is an umbrella term for impairments, activity limitations and participation restrictions in a biopsychosocial model. Disability can be assessed by means of questionnaires, clinical assessment (based on clinical history and physical examination), functional testing (for instance by a functional capacity evaluation), or by a combination of these methods. Measuring disability is important in rehabilitation research and practice because it can be used for quantifying current disability status, predicting future disability and evaluating treatment programs.

Questionnaires

A patient's self-report of perceived disability is often assessed with region-specific and general health questionnaires. Both should have good clinimetric qualities such as reproducibility, validity and responsiveness. The most frequently used neck disability questionnaires are the Neck Pain and Disability Scale (NPAD) and the Neck Disability Index (NDI), both translated into several languages. To be able to use NPAD and NDI these questionnaires must not only be translated but also culturally adapted and validated. These translations allow comparison of results of clinical research trials between countries. To investigate which questionnaire is most appropriate, clinimetric studies are needed where both questionnaires are applied simultaneously to the same sample of patients. The NPAD has not been formally translated into Dutch and consequently clinimetric qualities of the Dutch Language Version (NPAD-DLV) are unknown. The reliability of the NDI-DLV has been studied in patients with acute neck pain but not in patients with CNP other than patients with Whiplash Associated Disorder. The reproducibility, validity and responsiveness of the NDI-DLV in patients with CNP are unknown. Unlike the region-specific questionnaires, general health questionnaires cover physical functions as well as emotional and social functions. The Short Form-36 Health Survey (SF-36) is a widely used general health questionnaire assessing health in 8 domains. The construct validity of the SF-36 in patients with non-specific CNPin a tertiary outpatient rehabilitation setting is unknown.

Clinician assessment

An expert based assessment of disability by a clinician during physical examination on cervical mobility, pain provocation with cervical movements and behavioral signs provides relevant information about a patient's clinical status. Therefore, it is important to know the reliability of

these assessments and to explore the association with self-reported neck pain related disability and pain intensity. The physical dysfunction severity (PDS) scale has been used for the evaluation of physical impairment and treatment efficacy in controlled trials in primary care. PDS and perceived recovery were more sensitive to measure changes compared with self-reported disability and pain intensity. Interobserver reliability of the PDS has not been investigated and the construct validity is unknown.

Behavioral signs or 'non organic signs' (NOS) were developed in 1980 for patients with chronic back pain as screening tool to help identify patients who require more detailed psychosocial assessment. In 2000 a standardized set of cervical NOS (cNOS) was developed for patients with neck pain and cNOS were associated with prolonged disability and worker's compensation status. Prevalence of positive signs of cNOS as well as interobserver reliability for total scores of cNOS were not reported. Moreover, the construct validity of cNOS is unknown. The tests of cNOS were classified into five categories. The tests in the categories tenderness, regional disturbances and overreaction were based on lumbar NOS. Because of reasons described in chapter 6, we changed the tests in the categories simulation and distraction into the modified cNOS (mcNOS).

Functional Capacity Evaluation

A functional capacity evaluation (FCE) is a performance based assessment by an experienced physiotherapist with a standardized battery of tests to evaluate a patient's capacity to perform work-related activities. FCE's are used in work rehabilitation programs for measurement of disability, return to work recommendations and in medico-legal issues. However, no validated neck specific FCE has been described in peer reviewed literature.

In **chapter 1** the general introduction of this thesis was described. This thesis focused on clinimetric properties of instruments to measure disability in patients with non-specific CNP in tertiary rehabilitation setting. The research questions were:

- What is the test-retest reproducibility (test-retest reliability and agreement) of the Dutch language version of the NPAD and NDI?
- What is the content validity, internal consistency and construct validity of these questionnaires?
- What are relevant changes in total scores on the NPAD and NDI after a multidisciplinary outpatient pain rehabilitation and which questionnaire is most responsive to change?
- What is the construct validity of the SF-36?
- What is the interobserver reliability and what is the construct validity of the PDS?
- What is the interobserver reliability and what is the construct validity of the mcNOS?
- Can a content valid FCE be developed for patients with work-related neck pain?

The aim of the study described in **chapter 2** was to analyze the test-retest reliability and agreement of the NPAD and the NDI in a tertiary rehabilitation setting. The NPAD was translated from English into Dutch according to established guidelines. The study sample consisted of 34 patients (mean age 37.5 years, 68% female) with non-specific CNP admitted for multidisciplinary outpatient pain rehabilitation. The patients completed the questionnaires twice prior to treatment with a mean test-retest interval of 18 days. The intraclass correlation coefficient (ICC) of the NPAD was 0.76 (95% confidence interval (CI) 0.57–0.87) and of the NDI 0.84 (95% CI 0.69–0.92).The limits of agreement of the NPAD and NDI were respectively ± 20.9 (scale 0–100) and ± 6.5 (scale 0–50). It was concluded that the reliability of the NPAD and NDI was acceptable. The variation (instability) in the NPAD total scores was relatively large and larger than the variation of the NDI.

The aim of the study described in **chapter 3** was to analyze the validity of the NPAD and NDI in a tertiary rehabilitation setting. The study sample consisted of 112 patients (mean age 38.4 years, 70% female) with non-specific CNP admitted for multidisciplinary outpatient pain rehabilitation. The patients completed NPAD, NDI, SF-36, visual analog scale $(VAS)_{pain}$ and $VAS_{disability}$ prior to treatment. Twenty seven hypotheses were formulated regarding validity. NPAD and NDI were evaluated for content validity (normal distribution total scores, missing items, floor and ceiling effects), internal consistency (Cronbach's alpha and Spearman item-total correlations), construct validity (Pearson correlations with SF-36 domains, VAS_{pain} and $VAS_{disability}$ and Pearson correlation between total scores of NPAD and NDI). The study showed that NPAD and NDI scores were distributed normally and missing items were negligible. Floor and ceiling effects were absent in NPAD and in the NDI 2 items had floor effects and 1 item had a ceiling effect. Cronbach's alpha of the NPAD was 0.93 and of the NDI 0.83. Item-total correlations ranged from 0.45 to 0.73 for the NPAD and for the NDI from 0.40 to 0.64. The correlation between respectively the NPAD and NDI and: SF-36 domains ranged from -0.36 to -0.70 and from -0.34 to -0.63; VAS_{pain} was 0.54 and 0.43; $VAS_{disability}$ was 0.57 and 0.52. The correlation between the total scores of NPAD and NDI was 0.77. Twenty-six hypotheses were not rejected and one hypothesis was rejected. It was concluded that the NPAD and NDI were valid measures of self-reported neck-pain related disability.

The aims of the study described in **chapter 4** were to investigate relevant changes in total scores on the NPAD and NDI after a pain rehabilitation program and to determine which questionnaire is the most responsive in patients with non-specific CNP in a tertiary rehabilitation setting. The study sample consisted of 111 patients (mean age 38.7 years, 63% female) admitted for multidisciplinary outpatient pain rehabilitation. Patients completed NPAD and NDI prior to the start of the rehabilitation program and after completion of the program together with the global perceived effect (GPE) scale. After the start of the rehabilitation program 35 patients dropped out and 76 patients completed the program. Based on GPE scores patients were dichotomized into "improved" and "stable". To investigate relevant change the Minimal Detectable Change (MDC) on the one hand and the Minimal Important Change (MIC) with the receiver operator characteristic (ROC) cut-off point on the other were assessed. Comparison of responsiveness

was performed using areas under the ROC curve (AUC) and correlations between change scores of NPAD and NDI, and GPE. MDC and MIC on NPAD (scale 0–100) were 31.7 and 11.5 points respectively. MDC and MIC on NDI (scale 0–50) were 8.4 and 3.5 points respectively. Changes should exceed this MDC or MIC cut-off to be interpreted as relevant. AUC was 0.75 for both NPAD and NDI. Correlations between change scores of NPAD and NDI and GPE were, respectively 0.48 (95%) CI 0.29–0.64) and 0.49 (95% CI 0.30–0.64). It was concluded that relevant change on both NPAD and NDI assessed with MDC and MIC resulted in different cut-offs and consequently with different amounts of certainty that the patient had improved. Responsiveness of NPAD and NDI was similar.

The aim of the study described in **chapter 5** was to investigate the construct validity of the SF-36 in patients with non-specific CNP in a tertiary rehabilitation setting. The study sample consisted of 91 patients (mean age 38.7 years, 66% female) admitted for multidisciplinary outpatient pain rehabilitation. Patients completed the SF-36 prior to the start of the rehabilitation program. Sixteen a priori hypotheses were formulated regarding construct validity. SF-36 domain scores of patients with CNP were compared with general population reference values using standardized differences. For both the SF-36 physical and mental component summary (PCS and MCS) differences between primary and tertiary care setting, males and females, age groups, litigants and non-litigants, patients with and without workers compensation and with ≥3 versus ≤2 concomitant complaints were analyzed with independent t-tests. Difference between PCS and MCS score was analyzed with a paired t-test. All SF-36 domain scores were significantly lower than general population norm values. The domain scores 'role physical', 'bodily pain', 'vitality', 'social functioning' and 'role emotional' were relevantly (≥1 SD) lower. SF-36-PCS and SF-36-MCS scores were significantly lower than in a primary care setting. SF-36-PCS score was significantly lower for patients with workers compensation, and patients with ≥3 concomitant complaints. SF-36-MCS score was significantly lower for the age group ≥39 years. Twelve hypotheses were not rejected and 4 were rejected. It was concluded that the SF-36 has good construct validity and can be used to measure self-reported general health in patients with non-specific CNP within an outpatient tertiary rehabilitation setting.

The aims of the study described in **chapter 6** were to explore interobserver reliability of physical dysfunction severity (PDS) as a measure for physical impairment of the cervical spine and modified cervical 'non organic signs' (mcNOS) as a measure for behavioral signs, and to explore construct validity of PDS and mcNOS in a tertiary rehabilitation setting. The study sample consisted of 51 patients (mean age 38.5 years, 63% female) with non-specific CNP admitted for multidisciplinary outpatient pain rehabilitation. Two observers independently assessed PDS and mcNOS in these patients prior to treatment with a mean test-retest interval of 3 weeks. Interobserver reliability for total scores of PDS and mcNOS was expressed as intraclass correlation (ICC). Interobserver agreement for each 'non organic sign' was calculated as absolute agreement and Cohen's Kappa. Construct validity was expressed as Spearman correlation between PDS and mcNOS on the one hand, and Neck Pain and Disability Scale (NPAD) and Numeric Rating

Scale (NRS) for pain on the other hand. Interobserver reliability for PDS and mcNOS was ICC = 0.72 and ICC = 0.78. Agreement for 'non organic signs' ranged from 63% to 88%. Kappa values ranged from 0.14 to 0.54. Correlations of PDS and mcNOS with NPAD were respectively 0.26 and 0.49, and with NRS_{pain} 0.32 and 0.37. It was concluded that interobserver reliability of both PDS and mcNOS was acceptable. The interobserver agreement for the individual 'non organic signs' ranged from poor to acceptable. Construct validity of PDS and mcNOS appeared satisfactory.

The aim of the study described in **chapter 7** was to develop a Functional Capacity Evaluation (FCE) which is content valid for determining functional capacity in patients with work-related neck disorders (WRND). A review of literature was conducted to identify physical risk factors for WRND. Evidence was found that physical risk factors contribute in development of WRND. Physical risk factors were repetitive movements, forceful movements, awkward positions and static contractions of the neck or the neck/shoulder region. An FCE was designed based on the risk factors identified. To cover all risk factors eight tests were selected: repetitive side reaching, repetitive reaching overhead, static overhead work, front carry, forward static bend neck, overhead lift and the neck strength test. It was concluded that content validity of the neck-FCE was established by providing the rationale, specific objectives and operational definitions of the 8 tests of this FCE.

An overview and discussion of the mean results of this thesis are given in **chapter 8**. In measuring patients' self-report of perceived disability both NPAD and NDI have acceptable clinimetric properties but the NDI is slightly better. Although the SF-36 provides a more comprehensive measure of disability than NPAD and NDI the construct neck pain related disability is better reflected by NPAD and NDI. In the clinician-based assessment of disability by the clinician the PDS for measuring physical impairment and the mcNOS for measuring behavioural signs may offer relevant information in the biopsychosocial diagnosis of patients with CNP. For the performance based assessment of disability the proposed neck FCE may be useful. Generalizability of the results of the studies may be limited because the patient samples were largely consisting of patients with moderate neck pain related disability. Finally in this chapter the clinical implications of the results of the presented studies for measuring neck pain related disability are discussed and recommendations for further research are suggested.

Samenvatting

Nekpijn is een van de meest voorkomende klachten van het bewegingsapparaat. De 1-jaars prevalentie van chronische nekpijn (CNP) in de algemene bevolking is ongeveer 20%. Bij het merendeel van de patiënten met nekpijn kan geen specifieke onderliggende aandoening worden vastgesteld en wordt de nekpijn omschreven als 'aspecifieke' of 'mechanische' nekpijn. De term 'mechanische' nekpijn wordt wel gebruikt omdat de nekpijn wordt veroorzaakt of versterkt door bewegingen van de nek en door lang in dezelfde positie houden van nek en hoofd. Beperkingen in activiteiten en in dagelijkse bezigheden inclusief werk kunnen het gevolg zijn van nekpijn. Het merendeel van de totale kosten van nekpijn in Nederland zijn indirecte kosten samenhangend met ziekteverzuim en arbeidsongeschiktheid.

'Disability' is een paraplubegrip voor stoornissen in structuren en functies van het lichaam, beperking in activiteiten en participatieproblemen in een biopsychosociaal model. Volgens dit model wordt het functioneren van de patiënt beïnvloed door biomedische, psychologische en sociale factoren. Disability kan worden gemeten met behulp van vragenlijsten, klinisch onderzoek (gebaseerd op anamnese en lichamelijk onderzoek), functionele tests (bijvoorbeeld in de vorm van een Functional Capacity Evaluation) of een combinatie van deze methoden.

Het meten van disability is belangrijk in het wetenschappelijke onderzoek en de klinische praktijk van de revalidatiegeneeskunde omdat het kan worden gebruikt voor het kwantificeren van de actuele disability, voorspellen van de toekomstige disability en evaluatie van behandelprogramma's.

Vragenlijsten

De door patiënten ervaren disability wordt vaak bepaald met regio-specifieke en algemene gezondheidsvragenlijsten. Deze moeten beide goede klinimetrische eigenschappen hebben, zoals reproduceerbaarheid, validiteit en responsiviteit.

De meest gebruikte nekpijnvragenlijsten zijn de 'Neck Pain and Disability Scale' (NPAD) en de 'Neck Disability Index' (NDI). Om de NPAD en NDI te kunnen gebruiken, moeten ze niet alleen vertaald zijn, maar ook cultureel worden aangepast en gevalideerd. Dit maakt het mogelijk dat de resultaten van klinische trials, uitgevoerd in verschillende landen, kunnen worden vergeleken. Om te onderzoeken welke vragenlijst het meest geschikt is, zijn klinimetrische studies nodig waarbij beide vragenlijsten naast elkaar worden gebruikt bij dezelfde groep patiënten. De NPAD is nog niet vertaald in het Nederlands en dientengevolge zijn de klinimetrische eigenschappen van de Nederlandse versie onbekend. De betrouwbaarheid van de Nederlandse versie van de NDI is bestudeerd in patiënten met acute nekpijn, maar niet in patiënten met CNP, uitgezonderd patiënten met een whiplash. De reproduceerbaarheid, validiteit en responsiviteit van de NDI in patiënten met CNP zijn onbekend.

De algemene gezondheidsvragenlijsten bevatten – in tegenstelling tot de regio-specifieke – naast items over fysieke functies, ook uitgebreid items over emotionele en sociale functies. De Short Form-36 Health Survey (SF-36) is een veel gebruikte algemene gezondheidsvragenlijst

die de gezondheid in 8 domeinen meet. De constructvaliditeit van de SF-36 in patiënten met aspecifieke CNP in een universitaire revalidatiepolikliniek is onbekend.

Klinisch onderzoek

Voor de inschatting van de mate van disability door de clinicus levert lichamelijk onderzoek van de beweegbaarheid van de nek, provocatie van pijn tijdens nekbewegingen en ziektegedrag relevante informatie op. Daarom is het belangrijk de betrouwbaarheid te weten van deze onderzoeksmethoden en de samenhang na te gaan met de door de patiënt gerapporteerde nekpijngerelateerde disability en de intensiteit van de pijn.

De Physical Dysfunction Severity (PDS) bepaald door de clinicus, is in Nederland in eerder onderzoek gebruikt voor de evaluatie van fysieke stoornissen en behandelresultaten in gerandomiseerde studies in de eerste lijn. Hierbij was de PDS gevoeliger voor het vaststellen van behandeleffecten dan de mate van disability gemeten met de NDI en de pijnintensiteit. De interbeoordelaarsbetrouwbaarheid van de PDS is niet bestudeerd en de constructvaliditeit is onbekend.

Het meten van ziektegedrag of 'non-organic signs' (NOS) bij het lichamelijk onderzoek werd voor patiënten met chronische lage rugpijn als een screeningsinstrument ontwikkeld om te kunnen helpen bij het identificeren van patiënten bij wie verder onderzoek naar de invloed van psychosociale factoren is gewenst.

Een gestandaardiseerd onderzoek naar ziektegedrag (cNOS) bij het onderzoek van nekpatiënten werd eveneens ontwikkeld en het bleek dat de cNOS samenhing met beperkingen op langere termijn en ziekteverzuim. Hoe vaak positieve cNOS-tests aanwezig waren bij nekpatiënten, evenals de interbeoordelaars-betrouwbaarheid van de totaal score van cNOS, werd niet gerapporteerd. Bovendien is de constructvaliditeit van de cNOS onbekend. De tests van cNOS werden onderverdeeld in 5 categorieën. De tests in de categorieën 'gevoeligheid', 'regionale stoornissen' en 'overreactie' waren gebaseerd op de NOS bij patiënten met rugpijn. Vanwege redenen beschreven in Hoofdstuk 6, veranderden we de tests in de categorieën 'simulatie' en 'afleiding' in de gemodificeerde cNOS (mcNOS).

Functionele capaciteitsevaluatie

Een functionele capaciteitsevaluatie (FCE) is een testbatterij van functionele tests, afgenomen door een daarin getrainde fysiotherapeut, om de functionele mogelijkheden 'performance based' te meten bij het uitvoeren van werkgerelateerde activiteiten in een laboratoriumsetting. FCE's worden gebruikt in de arbeidsrevalidatie voor het vaststellen van beperkingen, aanbevelingen voor reïntegratie in werk en in medische/juridische zaken. Er is echter geen gevalideerde nekspecifieke FCE in de literatuur beschreven.

In **Hoofdstuk 1** werd de algemene introductie van dit proefschrift beschreven. Dit proefschrift richtte zich op de klinimetrische eigenschappen van methoden om disability te meten bij patiënten met aspecifieke CNP in een universitaire revalidatiepolikliniek. De volgende onderzoeksvragen werden geformuleerd:

- Wat is de test-hertestreproduceerbaarheid (test-hertestbetrouwbaarheid en overeenstemming) van de Nederlandse versie van de NPAD en NDI?
- Wat is de inhoudsvaliditeit, interne consistentie en constructvaliditeit van de NPAD en NDI?
- Wat zijn relevante veranderingen in de totaalscores op de NPAD en NDI na een multidisciplinair poliklinisch pijnrevalidatieprogramma en welke vragenlijst is het meest gevoelig om veranderingen vast te stellen?
- Wat is de constructvaliditeit van de SF-36?
- Wat is de interbeoordelaarsbetrouwbaarheid en de constructvaliditeit van de NDI?
- Wat is de interbeoordelaarsbetrouwbaarheid en de constructvaliditeit van de mcNOS?
- Kan een inhoudsvalide nek-FCE worden ontwikkeld om gebruikt te worden bij patiënten met werkgerelateerde nekpijn?

Het doel van de studie beschreven in **Hoofdstuk 2** was het bepalen van de test-hertestbetrouwbaarheid en overeenstemming (stabiliteit) van de NPAD en NDI bij patiënten van een universitaire revalidatiepolikliniek. De NPAD werd vanuit het Engels vertaald in het Nederlands, overeenkomstig de daarvoor geldende richtlijnen. De studie werd uitgevoerd bij 34 patiënten (gemiddelde leeftijd 37,5 jaar, 68% vrouwen) met aspecifieke CNP geïncludeerd voor een multidisciplinair poliklinisch pijnrevalidatie programma. De patiënten vulden de vragenlijsten twee keer in, met een gemiddeld test-hertestinterval van 18 dagen, voor de start van de behandeling. De Intraclass-correlatiecoëfficiënt (ICC) van de NPAD was 0,76 (95% betrouwbaarheidsinterval (CI) 0,57–0,87) en van de NDI 0,84 (95% CI 0,69–0,92). De 'limits of agreement´ van de NPAD en NDI waren respectievelijk + 20,9 (schaal 0–100) en + 6,5 (schaal 0–50). Geconcludeerd werd dat de betrouwbaarheid van de NPAD en NDI acceptabel is. De variatie (instabiliteit) in de totaalscores van de NPAD was relatief groot en groter dan de variatie van de NDI.

Het doel van de studie beschreven in **Hoofdstuk 3** was het bepalen van de validiteit van de NPAD en de NDI bij patiënten van een universitaire revalidatiepolikliniek. De studie werd uitgevoerd bij 112 patiënten (gemiddelde leeftijd 38,4 jaar, 70% vrouwen) met aspecifieke CNP geïncludeerd voor een multidisciplinair poliklinisch pijnrevalidatieprogramma. De patiënten vulden de NPAD, NDI, SF-36, visueel analoge schaal $(VAS)_{pijn}$ en de $VAS_{beperkingen}$ in voor de start van de behandeling. Zevenentwintig hypotheses werden geformuleerd met betrekking tot de validiteit. NPAD en NDI werden onderzocht op contentvaliditeit (normale verdeling van de totaalscores, gemiste items, vloer- en plafondeffecten), interne consistentie (Cronbach's alpha en Spearman

item-totaalscores-correlaties), constructvaliditeit (Pearson correlaties met SF-36 domeinscores, VAS_{pijn} en $VAS_{beperkingen}$ en Pearson correlatie tussen de totaalscores van NPAD en NDI). De studie toonde aan dat de NPAD en NDI totaalscores normaal waren verdeeld en dat het aantal niet ingevulde items te verwaarlozen was. Vloer- en plafondeffecten waren afwezig in de NPAD en in de NDI hadden 2 items vloereffecten en 1 item had een plafondeffect. Cronbach's alpha van de NPAD was 0,93 en van de NDI 0,83. Item-totaalscore-correlaties varieerden voor de NPAD van 0,45 tot 0,73 en voor de NDI van 0,40 tot 0,64. De correlaties tussen enerzijds de NPAD en de NDI en anderzijds: de SF-36 domeinscores varieerden van -0,36 tot -0,70 en van -0,34 tot -0,63; de VAS_{pijn} waren 0,54 en 0,43 en de $VAS_{beperkingen}$ waren 0,57 en 0,52. De correlatie tussen de totaalscores van NPAD en NDI was 0,77. Zesentwintig hypotheses werden niet verworpen en één hypothese werd verworpen. Geconcludeerd werd dat de NPAD en NDI valide meetinstrumenten zijn voor zelfgerapporteerde nekpijngerelateerde disability.

Het doel van de studie beschreven in **Hoofdstuk 4** was het bepalen van de relevante verandering in totaalscores van de NPAD en de NDI na een pijnrevalidatieprogramma en uit te maken welke vragenlijst het meest responsief is in patiënten met aspecifieke CNP in een universitaire revalidatiepolikliniek. De studie werd uitgevoerd bij 111 patiënten (gemiddelde leeftijd 38,7 jaar, 63% vrouwen) geïncludeerd voor een multidisciplinair poliklinisch pijnrevalidatieprogramma. De patiënten vulden de NPAD en NDI in voor de start van het revalidatieprogramma en na afloop van het programma met tevens de ervaren verandering in klachten (GPE). Na de start van het revalidatieprogramma vielen 35 patiënten uit en maakten 76 patiënten het programma af. Gebaseerd op de GPE werden deze 76 patiënten verdeeld in 'verbeterde' en 'stabiele' patiënten. Om de relevante verandering te bestuderen werden de minimal detectable change (MDC) en de minimal important change (MIC) volgens het ROC-curve afkappunt bepaald. De vergelijking van de responsiviteit werd uitgevoerd door gebruik te maken van de oppervlaktes onder de ROC-curves (de AUC's) en de correlaties tussen de veranderingen in de totaalscores van de NPAD en de NDI enerzijds en de GPE anderzijds. MDC en MIC op de NPAD (schaal0-100) waren respectievelijk 31,7 en 11,5 punten. MDC en MIC op de NDI (schaal 0–50) waren respectievelijk 8,4 en 3,5 punten. De veranderingen in de totaalscores zullen groter moeten zijn dan deze MDC of MIC afkappunten om als relevant te worden beschouwd. De AUC was 0,75 voor zowel de NPAD als de NDI. De correlaties tussen enerzijds de veranderingsscores van NPAD en NDI en anderzijds de GPE waren respectievelijk 0,48 (95% CI 0,29–0,64) en 0,49 (95% CI 0,38–0,64). Geconcludeerd werd dat relevante verandering op de NPAD en NDI bepaald met de MDC en de MIC resulteert in verschillende afkappunten en dientengevolge ook met verschillende mate van zekerheid dat de patiënt is verbeterd. Daarnaast was de responsiviteit van de NPAD en NDI gelijk.

Het doel van de studie beschreven in **Hoofdstuk 5** was het bepalen van de constructvaliditeit van de SF-36 in patiënten met aspecifieke CNP in een universitaire revalidatiepolikliniek. De studie werd uitgevoerd bij 91 patiënten (gemiddelde leeftijd 38,7, 66% vrouwen) geïncludeerd voor een multidisciplinair poliklinisch pijnrevalidatieprogramma. De patiënten vulden de SF-36 in voor de start van het revalidatieprogramma. Zestien hypotheses werden van te voren gefor-

muleerd met betrekking tot de constructvaliditeit. De SF-36 domeinscores van patiënten met CNP werden vergeleken met referentiewaarden van de algemene bevolking, gebruik makend van gestandaardiseerde verschillen. Voor zowel de SF-36 fysieke als mentale component somscore (PCS en MCS) werden de verschillen tussen patiënten uit de eerstelijns en derdelijns gezondheidszorg, mannen en vrouwen, leeftijdsgroepen, wel of geen lopende juridische procedure, wel of geen ziektewet en met >3 versus <2 bijkomende klachten bepaald met onafhankelijke t-tests. De verschillen tussen de PCS en MCS scores werden bepaald met een gepaarde t-test. Alle SF-36 domeinscores waren statistisch significant lager dan de referentiewaarden van de algemene bevolking. De domeinscores 'rol fysiek', 'lichaamspijn', 'vitaliteit', 'sociaal functioneren' en 'rol emotioneel' waren relevant (>1 SD) lager. De SF-36-PCS en SF-36-MCS scores waren statistisch significant lager dan in de eerstelijns gezondheidszorg. De SF-36-PCS score was statistisch significant lager voor patiënten in de ziektewet en voor patiënten met >3 bijkomende klachten. De SF-36 MCS score was statistisch significant lager voor de leeftijdsgroep >39 jaar. Twaalf hypotheses werden niet en 4 wel verworpen. Geconcludeerd werd dat de SF-36 een goede constructvaliditeit heeft en kan worden gebruikt om zelfgerapporteerde algehele gezondheid te meten in patiënten met aspecifieke CNP in een universitaire revalidatiepolikliniek.

Het doel van de studie beschreven in Hoofdstuk 6 was een exploratief onderzoek te verrichten naar de interbeoordelaarsbetrouwbaarheid van de PDS als maat voor de fysieke stoornis van de hals-wervelkolom en de mcNOS als maat voor ziektegedrag of 'non organic signs'. Daarnaast werd een exploratief onderzoek verricht naar de constructvaliditeit van de PDS en mcNOS in een universitaire revalidatiepolikliniek. De studie werd uitgevoerd bij 51 patiënten (gemiddelde leeftijd 38,5 jaar, 63% vrouwen) met aspecifieke CNP geïncludeerd voor een multidisciplinair poliklinisch pijnrevalidatieprogramma.

Twee beoordelaars bepaalden onafhankelijk van elkaar de PDS en mcNOS bij deze patiënten, voorafgaand aan de behandeling met een gemiddeld test-hertestinterval van 3 weken. De interbeoordelaarsbetrouwbaarheid voor de totaal scores van PDS en mcNOS werd uitgedrukt als intraclass correlatie (ICC). De interbeoordelaarsovereenstemming voor elk 'non organic sign' werd berekend als percentage absolute overeenstemming en op basis van Cohen´s kappa. De constructvaliditeit werd uitgedrukt als Spearman correlaties tussen enerzijds PDS en mcNOS en anderzijds NPAD en NRS_{pijn}. Voor de interbeoordelaarsbetrouwbaarheid van de PDS was de ICC 0,72 en van de mcNOS was de ICC 0,78. De overeenstemming van de 'non organic signs' varieerde van 63% tot 88% en de kappawaarden varieerden van 0,14 tot 0,54. De correlaties van de PDS en mcNOS met de NPAD waren respectievelijk 0,26 en 0,49 en met de NRS_{pijn} 0,32 en 0,37. Geconcludeerd werd dat de interbeoordelaarsbetrouwbaarheid van de PDS en mcNOS acceptabel is. De interbeoordelaarsovereenstemming voor de 'non organic signs' varieerde van zwak tot acceptabel. De constructvaliditeit van PDS en mcNOS bleek bevredigend te zijn.

Het doel van de studie beschreven in **Hoofdstuk 7** was een functionele capaciteitsevaluatie (FCE) te ontwikkelen die inhoudsvalide is voor de bepaling van de functionele capaciteit in

patiënten met werkgerelateerde nekaandoeningen (WRND). Een review werd uitgevoerd om de fysieke risicofactoren te identificeren van WRND. Er werd bewijs gevonden dat bepaalde fysieke risicofactoren bijdragen in de ontwikkeling van WRND. De factoren zijn: repeterende bewegingen, bewegingen waarbij kracht moet worden gezet, lastige werkhoudingen in extreme posities en statische duurcontracties van de spieren van de nek- of nek-schouderregio. Een FCE werd ontworpen gebaseerd op de geïdentificeerde risicofactoren. Om al de risicofactoren af te dekken werden 8 tests geselecteerd: repeterend links en rechts zijwaarts reiken, repeterend boven het hoofd reiken, statisch boven het hoofd werken, voor het lichaam dragen, zitten met statische belasting van de voorovergebogen nek, boven het hoofd tillen en de nekkrachttests. Geconcludeerd werd dat een inhoudsvalide nek-FCE kon worden ontwikkeld door de rationalisaties, specifieke doelen en operationele definities te geven van de 8 tests van deze FCE.

Een overzicht en discussie van de belangrijkste resultaten van dit proefschrift zijn beschreven in **Hoofdstuk** 8. Bij het meten van zelf gerapporteerde disability hebben zowel de NPAD als de NDI acceptabele klinimetrische eigenschapen, maar de NDI iets betere. Alhoewel de SF-36 een completere meting van de disability geeft dan de NPAD en de NDI, wordt het construct 'nekpijn gerelateerde disability' beter weergegeven door de NPAD en NDI. In de bepaling van de disability door de clinicus, kunnen de PDS -voor het meten van de fysieke stoornis- en de mcNOS -voor het meten van ziektegedrag- relevante informatie geven in de biopsychosociale diagnostiek van patiënten met CNP. Voor de performance based meting van disability kan de voorgestelde nek-FCE bruikbaar zijn. De generaliseerbaarheid van de resultaten van de studies wordt beperkt door het feit dat de patiëntengroepen grotendeels bestonden uit patiënten met matige nekpijn gerelateerde disability.Tenslotte werden in dit hoofdstuk de klinische implicaties van de resultaten van de gepresenteerde studies voor het meten van nekpijn gerelateerde disability besproken en werden er suggesties gedaan voor verder onderzoek.

EXPAND

Wetenschappelijk onderzoek afdeling Revalidatiegeneeskunde – Centrum voor Revalidatie UMCG

EXPAND

Extremities, Pain and Disability

Missie: EXPAND draagt bij aan participatie en kwaliteit van leven van mensen met aandoeningen en amputaties van de extremiteiten of met pijn aan het bewegingsapparaat.

EXPAND omvat twee speerpunten: onderzoek naar aandoeningen aan en amputaties van extremiteiten met nadruk op stoornissen, activiteiten en participatie en onderzoek naar chronische pijn en arbeidsparticipatie. EXPAND draagt bij aan het UMCG-brede thema Healthy Ageing.

Research Department of Rehabilitation Medicine – Center for Rehabilitation UMCG

EXPAND

Extremities, Pain and Disability

Mission: EXPAND contributes to participation and quality of life of people with conditions and amputations of the extremities and musculoskeletal pain.

EXPAND focuses on two spearheads: research on the conditions and amputations of the extremities with emphasis on body functions and structures, activities and participations, and chronic pain and work participation. EXPAND contributes to Healthy Aging, the focus of the UMCG.

Dankwoord

Al twee keer eerder heb ik een poging gedaan om een promotietraject te starten. De eerste keer ging het om een electromyografisch schouderonderzoek bij het anatomisch laboratorium van de RU Groningen. In dezelfde tijd echter startte ik ook mijn praktijk en werd ik onderwijscoördinator van de artsenleergangen van de Stichting Manuele Geneeskunde (SMG) te Eindhoven. Daardoor kwam het onderzoek in het gedrang.

De tweede keer betrof het een op te zetten onderzoek bij patiënten met heupklachten, aangaande het effect van fysiotherapie, manuele therapie en voortgezette behandeling door de huisarts. Dit plan ontstond na een interessante epidemiologiecursus in 1991 van de RU Limburg met onder anderen als enthousiaste docenten Bart Koes en Geert van der Heijden. Samen met Prof. Willem Eisma, hoogleraar revalidatiegeneeskunde, Prof. Dr. Betty Meyboom-De Jong, hoogleraar huisartsgeneeskunde, en Drs. Geert van der Heijden, gezondheidswetenschapper bij de vakgroep Epidemiologie van de RU Limburg, diende ik in 1994 een onderzoeksvoorstel in bij de Ziekenfondsraad. Helaas werd dit voorstel niet gehonoreerd.

Vlak voor het staken van mijn activiteiten als onderwijscoördinator bij de Artsenleergangen Manuele Geneeskunde van de SMG kwamen toch weer de kriebels en zie daar: drie maal is scheepsrecht. In juni 2004 schreef ik een voorstel gericht aan Prof. Dr. Jan Geertzen. Het waren roerige bestuurlijke tijden op Beatrixoord en het Academisch Ziekenhuis te Groningen en een en ander kwam daardoor niet van de grond. Gelukkig stapte Jan Geertzen in het voor jaar van 2006 tijdens een rondje hardlopen bij mij het pad op – ik was in de tuin aan het werk – om te vragen hoe het nu ging. De zaak werd vlot getrokken en samen met Michiel Reneman ging een en ander toch van start.

Nu aan het einde van een lange rit kan ik spreken van een leerzaam en zeer tijdsintensief traject, omdat daarnaast ook andere ballen in de lucht moesten worden gehouden. Desalniettemin ben ik zeer gelukkig het hele traject te hebben kunnen doorlopen dankzij de bijdragen van velen.

Allereerst bedank ik mijn eerste promotor Prof. Dr. Jan Geertzen. Beste Jan, jij hebt het onderzoek zoals beschreven, vlot getrokken in roerige Groningse revalidatietijden. Als aanhanger van resultaatvoetbal ben je doortastend, snel en recht door zee, hetgeen ik als zeer prettig heb ervaren. Ik heb bewondering voor je capaciteiten als manager.

Daarnaast dank ik mijn tweede promotor Prof. Dr. Michiel Reneman. Beste Michiel, ik ken je vanaf het moment dat je student was aan de Academie voor Fysiotherapie in Groningen, waar ik docent was. Nu waren de rollen omgedraaid, ik de leerling van wat Jan Geertzen de 'golden boy' van de pijnrevalidatie heeft genoemd. De samenwerking was al vanaf de Arbeidsexploratie-tijd op Beatrixoord zeer plezierig. Je hield me goed bij de wetenschappelijke les, bleef altijd optimistisch en je hebt een groot gevoel voor humor. We delen de interesses voor wijn, bergwandelen, overbelastingsblessures bij de wat oudere sportende mannen en het verbouwen van huizen.

Prof. Dr. Pieter Dijkstra als derde promotor dank ik voor zijn deskundige inbreng en zeer prettige samenwerking. Beste Pieter, vanaf het begin wist ik het zeker, jij moest ook in mijn promotie-

commissie, want ik schatte je capaciteiten hoog in. Nu, dit is volledig uitgekomen. Verschillende keren heb je de boel flink opgeschud en werden er nieuwe wegen ingeslagen. Je kijkt scherp naar teksten en weet een en ander tot de kern van de zaak terug te brengen. Jammergenoeg heb ik je nog niet aan het bergwandelen kunnen krijgen, maar fietsen is natuurlijk ook gezond.

Drs. drs. Grietje de Vries. Jarenlang heb je deel uitgemaakt van de auteurs. Ik prijs de dag dat ik je via Michiel leerde kennen. Zonder jouw hulp bij het verwerken van de data, het maken van tabellen en figuren, Ref Work, indienen van artikelen en je kritische blik en punctualiteit, had ik het niet gered. Thuis had ik het over 'mijn Grietje', of het 'meisje met het rode haar'. Je bent ondanks je dubbele drs.schap bescheiden, heel plezierig om mee samen te werken en altijd goed gehumeurd. Ik wens je als PhD-student een mooie wetenschappelijke carrière toe. Fijn dat je mij, samen met mijn dochter Annelies, als paranimf wilt bijstaan.

David Reesink, bedankt voor je belangrijke inbreng in Hoofdstuk 7. Ik heb het destijds erg leuk gevonden je samen met Michiel Reneman te begeleiden.

De leden van de beoordelingscommissie, Prof. Dr. Hans Arendzen, Prof. Dr. Sjoerd Bulstra en Prof. Dr. Bart Koes, dank ik voor het zorgvuldig lezen en beoordelen van dit proefschrift.

Hans, jou kende ik al als middelbare schoolvriend. Lochem-Goor was voor ons de derby op het hockeyveld. De HBS-tijd met onze vrienden en natuurlijk onmisbare vriendinnenclub zit vol zoete herinneringen. In Groningen ging dit verder. Zelfde studentenhuizen, Aegir en de eetclub. Ook op Beatrixoord kruisten onze wegen, toen ik via Cor Muskee bij jou, als medisch directeur van Beatrixoord, kwam te werken. Vanzelfsprekend kwam je in de commissie. Hopelijk zullen we nog lang van eetclub- en andere activiteiten kunnen blijven genieten.

Sjoerd, via een borrel in Paterswolde hebben we elkaar leren kennen. Ondanks dat we later een zeer goed gesprek hadden is het jammer genoeg niet gelukt om samen met andere afdelingen een Master-na-Master-opleiding Musculoskeletal Medicin in Groningen te realiseren. Heel prettig dat jij in de commissie wilde plaatsnemen.

Beste Bart, via de reeds genoemde cursus van de RU Limburg in 1991 werd ik enthousiast voor de wetenschap. Jouw promotieonderzoek betrof de eerste in Nederland uitgevoerde RCT betreffende de effectiviteit van manuele therapie en fysiotherapie bij patiënten met rug- en nekklachten. We zagen elkaar nog verschillende malen op symposia. Je bent voor de manuele therapie en de musculoskeletale geneeskunde van grote betekenis. Je enthousiaste reactie op de uitnodiging om plaats te nemen in de commissie heb ik zeer gewaardeerd.

Mijn bijzondere dank gaat ook uit naar de patiënten, zonder wie dit onderzoek niet mogelijk was geweest. Ik besef dat het invullen van vragenlijsten een aanslag kan zijn op het evenwicht in belasting en belastbaarheid.

Cor Muskee, via jou kwam ik in 1995 op Beatrixoord alsnog in een academisch circuit te werken. Voor mij was het een nieuwe uitdaging me in de pijnrevalidatie te bekwamen bij de nestor van

de pijnrevalidatie Noord Nederland. Onze langdurige, 25-jarige samenwerking in de manuele geneeskunde en de pijnrevalidatie heb ik als zeer plezierig ervaren. Als vast tandem gaven we niet alleen cursussen in Eindhoven bij de SMG, maar ook bij de Corvu, het GAK en het UWV over vooral aspecifieke klachten van het bewegingsapparaat. Je enthousiasme was aanstekelijk. In de oude revalidatiepolikliniek van Prof. Willem Eisma op het AZG maakten we vele foto's voor talloze onderwijsklappers. Gedurende 20 jaar verzorgden we samen jaarlijks twee dagen van de landelijke basiscursus 'Houding en Beweging' in Groningen voor de aanstaande revalidatieartsen.

In al die jaren dat ik me heb bezig gehouden met onderzoek en behandeling van klachten van het bewegingsapparaat zijn de talloze collega's uit de manuele therapie/geneeskunde waarmee ik heb samengewerkt van grote betekenis geweest, waarvoor mijn zeer hartelijke dank. Pittige discussies over nieuwe ontwikkelingen en in de ban doen van obsolete zaken, vulden vele SMG-docenten bijeenkomsten in Holten. Prettige bijkomstigheid was dat goede diners met mooie wijnen daar deel van uitmaakten.

Met het pijnteam op Beatrixoord hebben we de afgelopen 17 jaren ups en downs gekend. Gelukkig zijn we de afgelopen jaren in rustiger vaarwater gekomen. Ik kijk al met al met veel plezier terug en zal me de deskundige en gedreven inzet van alle leden van het pijnteam blijven herinneren. Zeer hartelijk dank hiervoor.

Ook de geweldige inzet van de secretaresses op het UMCG en Beatrixoord moet worden genoemd. Marleen Speller wordt in het bijzonder bedankt voor haar medewerking aan het onderzoek. De collega's uit de medische staf van het Centrum voor Revalidatie waarmee ik zeventien jaar mocht samenwerken prijs ik om hun deskundigheid en grote betrokkenheid in de patiëntenzorg en hun grote inzet bij het onderwijs. Het was een groot plezier met zo'n positieve club te mogen werken. Mijn directe collega's Rita Schiphorst Preuper, Maikel Siebrecht en Jitze Posthumus dank ik voor de zeer plezierige samenwerking en hun stimulerende belangstelling voor mijn onderzoek. Het ga jullie goed!

Theo van Berkel, mijn studietijd-vriend, Indische buurt-genoot, Aegirpupil, fietsvriend en eet-clubgenoot wordt hartelijk bedankt voor zijn correcties op de Engelse tekst in Hoofdstukken 1, 6 en 8 en de Summary.

Renate Siebes heeft door haar zorgvuldige werk voor het drukrijp maken van het manuscript vele zorgen van mij kunnen wegnemen. Je bent net zo'n Pietje precies als je vader Koos Siebes is bij het inpakken van zijn rugzak. Zeer hartelijk dank.

Als je een lieve nicht hebt en haar man is kunstenaar, fotograaf en alleskunner, krijg je een mooi ontwerp voor de omslag van het boekje. Rutger Hiemstra en Hidda Schillhorn van Veen zeer bedankt.

Als laatste wil ik mijn naasten in het dankwoord betrekken. Ik draag het boekje mede op ter nagedachtenis aan mijn vader, die ik zeer heb bewonderd. Heel fijn dat mijn moeder dit nog kan meemaken. Lieve Annelies, Hans en Reinder, bedankt voor jullie niet aflatende belangstelling en

hand- en spandiensten. Lieke en Rutger wil ik hier ook graag bij betrekken. Annelies, omdat jij het klappen van de zweep ten aanzien van promoveren en de daarbij behorende nevenactiviteiten kent, vind ik het heel fijn dat je paranimf wil zijn.

Tot slot. Lieve Lideke, bedankt voor al je typewerk, correcties, eerste beoordelingen, organisatietalent, geduld en belangrijke familieondersteunende activiteiten. Fijn dat we straks meer vrije tijd zullen hebben om samen leuke dingen te doen.

About the author

CURRICULUM VITAE

Wim Jorritsma werd op 10 december 1947 geboren te Lochem. Hij behaalde het diploma van de Rijks HBS in 1965 en begon in hetzelfde jaar met de studie Geneeskunde aan de Rijksuniversiteit in Groningen. In zijn studententijd was hij zeer actief als trainer-coach bij de studentenroeivereniging Aegir en had hij veel interesse in de sportgeneeskunde. Tijdens zijn wachttijd voor bepaalde co-schappen begon hij in 1975 als parttime docent aan de Academie voor Fysiotherapie te Leeuwarden. In 1977 behaalde hij zijn artsexamen.Door zijn interesse in het bewegingsapparaat ging hij manuele therapie op de School voor Manuele Therapie in Utrecht studeren en haalde daar zijn diploma in 1980. In 1984 haalde hij zijn diploma in de manuele geneeskunde bij de Stichting Manuele Geneeskunde (SMG) in Eindhoven. Sinds 1984 voert hij parttime zijn praktijk manuele geneeskunde in Paterswolde. In 1994 haalde hij het Credentialing Examination van het McKenzie Institute Benelux te Roermond.

Van 1975 tot 1995 was hij docent op de Academies voor Fysiotherapie te Leeuwarden, Deventer en Groningen in de functionele en klinische anatomie, kinesiologie, anatomie in vivo en orthopedie. Als docent manuele therapie/manuele geneeskunde was hij sinds 1981 actief, eerst op de School voor Manuele Therapie in Utrecht en vanaf 1984 bij de Artsenleergangen Manuele Geneeskunde van de SMG in Eindhoven. Van 1985 tot 2005 was hij tevens onderwijscoördinator van deze stichting. Vanaf 1985 is hij docent en 'ärtzliche Leiter' van de opleiding Manuelle Therapie van het Verband Physikalische Therapie, Landesgruppe Nordrhein-Westfalen in Recklinghausen. Daarnaast gaf hij van 1985 tot 2010 regelmatig cursussen in onderzoek en kliniek van het bewegingsapparaat aan revalidatieartsen, bedrijfsartsen, verzekeringsartsen en huisartsen. In 2010 was hij medeoprichter van de Stichting Wetenschap & Onderwijs in de Musculoskeletal Medicine te Utrecht.

Vanaf 1995 tot december 2012 was hij voor 50% werkzaam in het Pijnteam van het revalidatiecentrum Beatrixoord te Haren en later ook op de polikliniek van het Centrum voor Revalidatie van het UMCG te Groningen. In die functie was hij sinds 2000 achtereenvolgens ook werkzaam bij Arbeidsexploratie Beatrixoord, in het Academisch Centrum voor Arbeid en Gezondheid van het Academisch Ziekenhuis te Groningen, later het UMCG, en vanaf 2008 tot heden in Team Arbeid in het Centrum voor Revalidatie UMCG, locatie Beatrixoord voor de multidisciplinaire Quick Scans en expertises.

PUBLICATIONS

International

Jorritsma W, Dijkstra PU, de Vries GE, Geertzen JH, Reneman MF (2012) Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index. Eur Spine J 21:2550–2557

Jorritsma W, de Vries GE, Dijkstra PU, Geertzen JH, Reneman MF (2012) Neck Pain and Disability Scale and Neck Disability Index: validity of Dutch Language Versions. Eur Spine J 21:93–100

Jorritsma W, de Vries GE, Geertzen JH, Dijkstra PU, Reneman MF (2010) Neck Pain and Disability Scale and the Neck Disability Index: Reproducibility of the Dutch Language Versions. Eur Spine J 19(10):1695–1701

Reesink DD, **Jorritsma W**, Reneman MF (2007) Basis for a functional capacity evaluation methodology for patients with work-related neck pain. J Occup Rehabil 17(3):436–449

Kuijer W, Brouwer S, Dijkstra PU, **Jorritsma W**, Groothoff JW, Geertzen JH (2005) Responsiveness of the Roland-Morris Disability Questionnaire: consequences of using different external criteria. Clin Rehabil 19(5):488–495

Reneman MF, **Jorritsma** W, Dijkstra SJ, Dijkstra PU (2003) Relationship between kinesiophobia and performance in a functional capacity evaluation. J Occup Rehabil 13(4):277–285

Reneman MF, **Jorritsma** W, Schellekens JM, Göeken LN (2002) Concurrent validity of questionnaire and performance-based disability measurements in patients with non specific low back pain. J Occup Rehabil 12(3):119–129

Reneman MF, Dijkstra SJ, **Jorritsma** W, Muskee C, Schiphorst Preuper HR, Göeken LN (2001) Assessment and treatment of chronic work-related pain disorders in an outpatient university rehabilitation setting in the Netherlands. Work 16(1):23–30

Winters JC, **Jorritsma** W, Groenier KH, Sobel JS, Meyboom-de Jong B, Arendzen HJ (1999) Treatment of shoulder complaints in general practice: long term results of a randomised single blind study comparing physiotherapy, manipulation and corticosteroid injection. BMJ 318(7195):1395–1396

National

Dikkeboer B, **Jorritsma W** (1983) Kinesiologie. Een handleiding bij het kinesiologie-onderwijs aan Academies voor Lichamelijke Opvoeding en Academies voor Fysiotherapie. Bohn, Scheltema & Holkema, Utrecht/Antwerpen. ISBN 9031305731.